

Non-Vertex-Balanced Factors in Random Graphs

Stefanie Gerke* Andrew McDowell*[†]

April 11, 2013

Abstract

We prove part of a conjecture by Johansson, Kahn and Vu [6] regarding threshold functions for the existence of an H -factor in a random graph $G(n, p)$. We prove that the conjectured threshold function is correct for any graph H which is not covered by its densest subgraphs. We also demonstrate that the main result of [6] generalises to multigraphs, digraphs, and a multipartite model.

Keywords: Random Graphs, Factors, Digraphs

1 Introduction

We will properly state our theorems later in the introduction, after introducing necessary notation and background. However, for readers already familiar with the background or willing to momentarily gloss over the details, let us immediately sketch our main results and methods. In a recent breakthrough (winning a 2012 Fulkerson Prize), Johansson, Kahn, and Vu [6] determined the threshold for a random graph G to be factorable by a strictly balanced fixed graph H , and they conjectured the threshold for every H . Our main result, Theorem 1.1, establishes their conjecture for ‘non-vertex-balanced’ graphs H , a class of graphs disjoint from strictly balanced ones. The positive side is the difficult one, and the main idea of the proof is, to cover a fraction of G with copies of a densest subgraph of H , then contract that subgraph to a point, extend the cover, and repeat. However, after the first

*Mathematics Department, Royal Holloway College, University of London, Egham TW20 0EX, England.

[†]e-mail `Andrew.McDowell.2010@live.rhul.ac.uk`

step, two things have changed: the graph H may have become a multigraph and, more significantly, we have committed to correspondences between some vertices of G and H . We manage these difficulties through Theorem 1.2, asserting that, if the vertices of the random graph G are partitioned into classes corresponding to vertices of H , then G almost surely has an H -factor which respects the partitioning. Our proof follows the steps of the proof in [6]; it is not especially inventive, but neither is it easy. Krivelevich [8] needed a special case and verified it, but [8] does not include the proof. The result is clearly useful, and it is therefore worth writing down the proof details. We prove a similar result for directed graphs.

Formally, for graphs H and G , an H -factor of G is a collection of vertex-disjoint copies of H in G that form a vertex cover of G . Clearly G can only contain an H -factor if $|V(H)|$ divides $|V(G)|$. We are mainly interested in random graphs on n vertices and we assume throughout the paper that $|V(H)|$ divides n .

We call a function $f(n)$ a *threshold* for a graph property K if, for an Erdős-Rényi random graph $G(n, p)$, (that is, the graph on n vertices where each edge is present with probability p independently of the absence or presence of any other edge)

$$\Pr(G(n, p) \text{ satisfies } K) \rightarrow \begin{cases} 1 & \text{if } p(n) = \omega(f(n)), \text{ and} \\ 0 & \text{if } p(n) = o(f(n)). \end{cases}$$

Since containing an H -factor is an increasing property (that is, adding edges does not destroy any H -factor) it is well known that a threshold function exists, see for example [5]. Note that a threshold is unique up to multiplicative positive constants so we will use Θ notation and with slight abuse of language we will speak of “the” threshold. The study of thresholds for various classes of graphs H has attracted considerable interest. The distinctions center around density properties of H . We define the *density* of a graph H on at least two vertices, as

$$d(H) = \frac{|E(H)|}{|V(H)| - 1}.$$

Let $m(H)$ be the maximum density of any subgraph of H , that is,

$$m(H) = \max \{d(H') : H' \subseteq H, |H'| \geq 2\}.$$

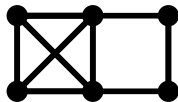
A graph H is called *balanced* if $m(H) = d(H)$, i.e., if no subgraph of H has density greater than that of H , and *strictly balanced* if every proper subgraph of H has density smaller than that of H .

For any vertex v of H , define the *local density* at v to be the maximum density restricted to subgraphs containing v ,

$$m(v, H) = \max \{d(H') : H' \subseteq H, |H'| \geq 2, v \in V(H')\}.$$

A graph H is *vertex balanced* if, for all $v \in H$, $m(v, H) = m(H)$.

Note that if H is balanced then it is vertex balanced: a densest subgraph of H is H itself, so $m(v, H)$ and $m(H)$ are both given by $H' = H$, for $m(v, H) = m(H) = d(H)$. Taking the contrapositive, if H is non-vertex balanced then it is not balanced, and not strictly balanced. Graphs may thus be partitioned into those that are non-vertex balanced, those that are strictly balanced, and the rest (those that are vertex balanced but not strictly balanced). An example of a non-vertex balanced graph is shown below.



The thresholds for H -factors for various fixed graph H have been of interest for a long time. The case $H = K_2$ is simply the threshold for G to have a perfect matching which has been known since 1966 [3], see also [2] for a more precise result. The next H -factor threshold result was for trees by Łuczak and Ruciński [9]. Note that matchings and trees are vertex-balanced. For sub-classes of non-vertex-balanced graphs, the threshold is known for graphs H whose minimum degree is less than $m(H)$ [1, 5]. In 2008 the seminal paper by Johansson, Kahn, and Vu [6] determined the threshold for all strictly balanced graphs (also resolving the so-called ‘Shamir’s problem’ on hypergraph matchings). The special case of finding the threshold of an H -factor for the strictly balanced graph $H = K_3$ had been described by Janson, Łuczak and Ruciński as one of the two ‘most challenging, unsolved problems in the theory of random structures’ [5, p. 96] and was first posed by Ruciński in 1992 [11] (the second problem was ‘Shamir’s problem’).

In their paper Johansson, Kahn and Vu conjecture thresholds for all graphs H , depending on whether H is vertex-balanced or not. We restate this formally as Conjecture 1 in Section 2. Our first main result establishes this conjecture for all graphs in the second category. More precisely, let $\text{th}_H(n)$ be the threshold function for $G(n, p)$ to contain an H -factor. We prove the following.

Theorem 1.1. *If H is non-vertex-balanced,*

$$\text{th}_H(n) = \Theta \left(n^{-1/m(H)} \right).$$

The main idea of the proof is, first, to embed the dense subgraphs of H , giving a ‘partial factor’ covering a corresponding proportion of the vertices of $G(n, p)$. We then collapse each such subgraph of H to a single vertex, giving a less dense strictly balanced graph (or possibly multigraph). Finally, we extend the partial factor to a full factor using Theorem 1.2, a generalisation to partitioned multigraphs of the strictly balanced result of [6].

To state Theorem 1.2 we need some more notation. Let $e_H = |E(H)|$, $v_H = |V(H)|$ and $V(H) = \{x_1, x_2, \dots, x_{v_H}\}$. Define the r -fold blowup $B(H, r)$ of H as an v_H -partite graph with parts V_1, V_2, \dots, V_{v_H} , each of size r , with an edge between $v_i \in V_i$ and $v_j \in V_j$ iff there is an edge between x_i and x_j in H . In a slight abuse of notation, let $H(n, p)$ be the random subgraph of $B(H, n/v_H)$ obtained by retaining each edge with probability p . Likewise, given a multigraph \mathcal{H} , we define the random multigraph $\mathcal{H}(n, p)$: the blowup $B(\mathcal{H}, n/v_{\mathcal{H}})$ has as many edges between $v_i \in V_i$ and $v_j \in V_j$ as there are edges between x_i and x_j in \mathcal{H} , and again $\mathcal{H}(n, p)$ is the random subgraph of $B(\mathcal{H}, n/v_{\mathcal{H}})$ obtained by retaining each edge with probability p .

The setup suggests looking for a ‘restricted’ \mathcal{H} -factor of $\mathcal{H}(n, p)$ where, for each copy of \mathcal{H} , each vertex belongs to the corresponding part of $\mathcal{H}(n, p)$. We will not in fact assume this, but the following theorem shows that below some threshold there is no factor, while above the threshold there is a factor of the restricted form.

Theorem 1.2. *Fix a multigraph \mathcal{H} (which may be a simple graph H). If \mathcal{H} is strictly balanced, then the threshold for $\mathcal{H}(n, p)$ to contain an \mathcal{H} -factor is*

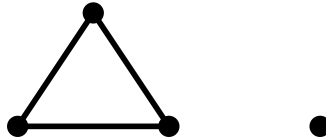
$$\text{th}_{\mathcal{H}}(n) = \Theta \left(n^{-1/m(\mathcal{H})} (\log n)^{1/|E(\mathcal{H})|} \right),$$

while if \mathcal{H} is not strictly balanced, the threshold satisfies

$$\text{th}_{\mathcal{H}}(n) = \mathcal{O}(n^{-1/m(\mathcal{H})+o(1)}).$$

In both cases, above the threshold there is w.h.p. an \mathcal{H} -factor in which, in every copy of \mathcal{H} , each vertex is in the corresponding part of $\mathcal{H}(n, p)$.

We note that there is a key difference between the partitioned and usual $G(n, p)$ thresholds for non-strictly balanced graphs. In $G(n, p)$, we show that for non-vertex balanced H , the $o(1)$ term can be completely eliminated, while it remains in the form of a log term for strictly balanced H . In the partitioned random graph however, there will always be a log term. This can be seen by considering the graph consisting of a triangle and a single isolated vertex.



In $G(n, p)$, this graph is easy to embed, as it is equivalent to a partial factor of triangles, taken over the whole graph, but only covering $3/4$ of the vertices. At this point, the remaining spare vertices immediately complete the factor. In the partitioned case, by fixing the position of these triangles, we are implying the existence of a full triangle factor over those corresponding partitions, and as such, by the result for strictly balanced graphs, we will require a log term, corresponding to the densest subgraphs.

Lastly we prove that the threshold for digraph factors coincides with that for graphs which is an easy consequence of Theorem 1.2. We define the random directed graph $D(n, p)$ with vertex set V of size n , such that for each pair of vertices u and v in V , there is an edge between them with probability p independently of all other edges, and each such edge is either (u, v) or (v, u) , with probability half each. We can prove the threshold for both strictly balanced and non-vertex-balanced digraphs, in $D(n, p)$. Note that for strictly balanced graphs we can also prove the partitioned form, i.e., the digraph form of Theorem 1.2 also holds.

Theorem 1.3. *Fix a digraph H . If H is strictly balanced, then the threshold function $\text{th}_H(n)$, for the random directed graph $D(n, p)$ to contain an H -factor is*

$$\text{th}_H(n) = \Theta \left(n^{-1/m(H)} (\log n)^{1/e_H} \right),$$

while if H is non-vertex-balanced,

$$\text{th}_H(n) = \Theta \left(n^{-1/m(H)} \right).$$

2 Preliminaries

We first note the following known results which we will need later.

Theorem 2.1. *(Ruciński [11]) Let H be a graph with at least one edge, and $F_H(\varepsilon, n)$ be threshold function for the property that $G(n, p)$ contains a partial H -factor covering all but at most εn vertices. Then for any fixed $\varepsilon > 0$ the threshold function satisfies,*

$$F_H(\varepsilon, n) = \Theta \left(n^{-1/m(H)} \right).$$

Theorem 2.2. *(Alon, Yuster [1]) Let H be a graph with minimum degree $\delta(H)$, satisfying $\delta(H) < m(H)$. Then*

$$\text{th}_H(n) = \Theta \left(n^{-1/m(H)} \right).$$

In their respective papers, stronger results than what is stated above are actually proved, but we use threshold notation for consistency.

The following two results can be found as Theorem 2.1 and 2.2 in [6]. This paper's aim is to provide a generalisation of the first result, which allows for improved bounds on the second.

Theorem 2.3. [6] *Let H be a strictly balanced graph with e_H edges. Then the threshold function, $\text{th}_H(n)$ for $G(n, p)$ to contain an H -factor satisfies*

$$\text{th}_H(n) = \Theta(n^{-1/d(H)}(\log n)^{1/e_H}).$$

Theorem 2.4. [6] *For H , an arbitrary fixed graph, the threshold function, $\text{th}_H(n)$ for $G(n, p)$ to contain a H -factor satisfies*

$$\text{th}_H(n) = \mathcal{O}(n^{-1/m(H)+o(1)}).$$

In [6], the authors define threshold functions $\text{th}_H^{[1]}(n)$ and $\text{th}_H^{[2]}(n)$, for a given fixed graph H . Firstly, $\text{th}_H^{[1]}(n)$ is defined as the threshold for every vertex in G to be covered by at least one copy of H , while $\text{th}_H^{[2]}(n)$ is the threshold for the property of satisfying the following two conditions:

1. every vertex of G is covered by at least one copy of H , and
2. for each $x \in V(H)$, there are at least n/v_H vertices x' of G for which some isomorphism of H takes x to x' .

This threshold is clearly a lower bound for the threshold for finding a factor $\text{th}_H(n)$ and the paper conjectures that they are, in fact, equal, and they prove this for strictly balanced H . In this paper, we will show this conjecture also holds for a large class of non-strictly balanced graphs.

The threshold $\text{th}_H^{[2]}(n)$ is completely determined and stated without proof in [6], for all graphs. For completeness we have included a proof below. Let $s_v = \min\{e(H') : H' \subseteq H, v \in V(H'), d(H') = m(v, H)\}$ and let s be the maximum over all s_v . Clearly $m(v, H) \leq m(G)$ for all v , with equality for at least one v . We are now ready to state and prove the following:

Lemma 2.5. *If for all $v \in V(H)$, $m(v, H) = m(H)$ then*

$$\text{th}_H^{[2]}(n) = \Theta(n^{-1/m(H)}(\log(n))^{1/s}).$$

Otherwise

$$\text{th}_H^{[2]}(n) = \Theta(n^{-1/m(H)}).$$

Proof. We clearly have 4 cases to consider. Firstly we will look at vertex-balanced graphs, i.e. those that satisfy, $m(v, H) = m(H)$ for all $v \in V(H)$. Condition 1 of $\text{th}_H^{[2]}(n)$, namely that each vertex of G is covered by at least one copy of H , is well studied and exact thresholds can be found as Theorem 3.22 in [5] and follow from results proved by Spencer in [13], and, in this case they are equal to our required bound.

Now we simply have to prove that condition 2 is also satisfied for $p = \omega(n^{-1/m(H)}(\log(n))^{1/s})$. We note that we clearly have that $Cp = \omega(n^{-1/m(H)})$ for any constant C .

Let $V(H) = \{1, 2, \dots, v_H\}$, our result will follow from partitioning the edge set of $G(n, p)$ into the union of random graphs G_0, G_1, \dots, G_{v_H} with edge probability p' , where $1 - p = \prod_{i=0}^{v_H} (1 - p') = (1 - p')^{v_H+1}$ and repeatedly applying Theorem 2.1 to find partial factors of H . We first apply it in G_0 , which has edge probability $p' > p/(v_H + 1) = \omega(n^{-1/m(H)})$, which is sufficient to apply 2.1 with $\varepsilon = 1/4$. This gives us a partial H -factor covering $3n/4$ of the vertices of $G(n, p)$ and hence $(1 - \varepsilon)n/v_H = 3n/(4v_H)$ vertices of $G(n, p)$ are covered by each vertex of H with high probability.

For each $i = \{1, 2, \dots, v_H\}$ we consider vertex i of H and the vertices of G_0 , that we have already covered by copies of i , and then, the random graph induced by the edges of G_i on the vertex set of G_0 , without those already covered vertices. This leaves us with a set of $n' = (1 - 3/(4v_H))n$ vertices in each G_i , that have not already been covered by a copy of the vertex i of H , with an independent random edge set. We can consider this as equivalent to the random graph $G(n', p')$, where $p' > p/(v_H + 1) = \omega((n')^{-1/m(H)})$ (assuming $|v_H| > 2$). This allows us to again, apply Theorem 2.1 to find another set of partial factors on $3/4$'s of the remaining vertices, giving us in total $(6/4v_H - 9/16v_H^2)n > n/v_H$, for $v_H > 1$, vertices covered by vertex i of H as required.

We now consider graphs that are non-vertex-balanced and so do not satisfy $m(v, H) = m(H)$ for all $v \in V(H)$. As before, the threshold for covering is known, and is in fact lower than our required threshold here.

The same argument for proving condition 2 as above applies since we only required $p = \omega(n^{-1/m(H)})$, so it follows that, for these graphs, both conditions are satisfied for $p = \omega(n^{-1/m(H)})$. It only remains to show that the threshold is not lower than this for such H . This follows from another result, proved by Ruciński and Vince [12]. They prove, that for any vertex of $G(n, p)$, the threshold for it being covered by a particular vertex v of H is $n^{-1/m(v, H)}$.

With the result above in mind, we define the following; for a vertex $v_G \in V(G(n, p))$, we let X_{v_G} be the indicator variable for v_G being covered by a copy of v , where $v \in V(H)$ satisfies $m(v, H) = m(H)$, namely $X_{v_G} = 0$

if it is not covered, and $X_{v_G} = 1$ if it is. Suppose that condition 2 is satisfied with high probability. Therefore we have that

$$\mathbb{E}\left(\sum_{v_G \in G(n,p)} X_{v_G}\right) = \sum_{v_G \in G(n,p)} \mathbb{E}(X_{v_G}) > n/v_H.$$

Suppose that $p = o(n^{-1/m(v,H)})$. We know that for p in this range, $X_{v_G} = 0$, with high probability, and therefore $\mathbb{E}(X_{v_G}) = o(1)$. Since there are only n such choices of v_G , we have a contradiction. Therefore the threshold $\text{th}_H^{[2]}(n)$ is not $o(n^{-1/m(H)})$, and so must be $n^{-1/m(H)}$, as required. \square

In [6], it is conjectured that $\text{th}_H^{[2]}(n) = \text{th}_H(n)$, so in light of the above, this can be restated as the following

Conjecture 1. [6] *If for all $v \in V(H)$, $m(v, H) = m(H)$, then*

$$\text{th}_H(n) = n^{-1/m(H)}(\log(n))^{1/s}.$$

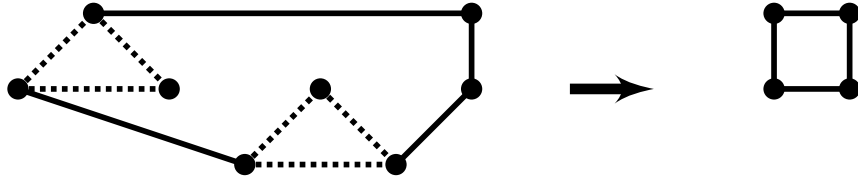
Otherwise

$$\text{th}_H(n) = n^{-1/m(H)}.$$

3 Theorem 1.1

The first case of Conjecture 1 has been proved for strictly balanced H , and now we will prove the second statement in its entirety, namely we prove that the threshold for containing an H -factor is $\text{th}_H(n) = \text{th}_H^{[2]}(n) = n^{-1/m(H)}$ for graphs where $m(v, H) < m(H)$ for some $v \in V(H)$. We begin by demonstrating that the result follows, assuming Theorem 1.2, and then in Section 4, we return to prove it.

In general terms, the main idea of this paper, is to ‘collapse’ dense subgraphs of H to get a new graph (or possibly multigraph) \mathcal{H} , which we will formally define later.

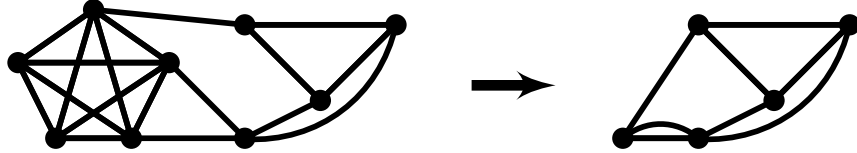


Since, we have $m(v, H) < m(H)$ for some v , we know that at least one vertex of H does not belong to any dense subgraphs of H . As a result, we will only need to cover a linear fraction of the vertices of $G(n, p)$ with these

dense subgraphs, since our factor will contain at least n/v_H vertices to be covered by copies of these less dense vertices.

Once we have embedded the dense sub-graphs, we then use Theorem 1.2, treating these (collapsed in H) embedded graphs as single vertices and finding a new, equally or less dense, \mathcal{H} -factor on these collapsed vertices, along with the remaining uncovered vertices of $G(n, p)$. This will translate to the required factor in our original graph.

To do this for the graph above, we would simply require a generalisation that allows us to partition our vertices and choose which vertex of \mathcal{H} will ‘cover’ the vertices of $G(n, p)$ in our factor. However, in a more general case, after collapsing vertices in H we may no longer be left with a graph, but a multigraph, hence the required level of generalisation to use this method.



In this example, the densest subgraph is clearly the K_5 , so we would collapse this to a single vertex. However, one vertex of H contains edges to two vertices of this subgraph, leaving us with a multigraph. It is also worth noting that the density of K_5 is 2.5 and since every vertex has degree at least 3, this is an example of a non-balanced graph, that could not be solved by the minimum degree result [1], and where we can provide the optimal threshold, improved on that provided by Theorem 2.4.

To define our collapsing method formally, we begin with some observations on the effects of vertex collapsing on the density of H . We know that $m(v, H) = m(H)$ for some vertices v , and these are the vertices that we collapse. For each such v , we, in turn, choose a subgraph H' such that $v \in H'$ and $d(H') = m(H)$. We now collapse all the vertices in H' into a single vertex. Giving us a new (possibly multi) graph, which we will call H_1 , which has the vertices of $H \setminus H'$, with an additional vertex v_1 , and an edge for each edge of H with an endpoint in $H \setminus H'$. We continue this process, going from H_i to H_{i+1} , at each stage, collapsing a subgraph of density $m(H)$ until none remain. The final graph which contains no subgraphs of density $m(H)$, we will call \mathcal{H} . We prove the rigour of this statement in the following lemma.

Lemma 3.1. *The collapsing process, described above, terminates after a finite number of steps, producing a unique multigraph \mathcal{H} , with $m(\mathcal{H}) < m(H)$.*

Proof. Firstly note that the density of H_1 , defined in the same way for multi-graphs as for graphs, is

$$d(H_1) = \frac{e(H_1)}{(v(H_1) - 1)} = \frac{e(H) - e(H')}{v(H) - v(H') + 1 - 1} = \frac{e(H) - e(H')}{(v(H) - 1) - (v(H') - 1)}. \quad (1)$$

Noting that $d(H) \leq m(H) = d(H')$, we can see that the above gives us, $d(H_1) \leq d(H) \leq m(H)$. If instead of H and H_1 , we consider any subgraph of H containing the vertices we are going to collapse and the resulting subgraph of H_1 , the same inequality shows that we have not created any subgraph in H_1 of density greater than $m(H)$. In fact, considering the following for positive numbers a, b, c and d ;

$$\frac{a - c}{b - d} \geq \frac{a}{b} \iff \frac{a}{b} \geq \frac{c}{d}$$

(assuming $b > d$), and noting that we only have equality on one side if we have it on both, it follows from (1) that any vertex that is in a subgraph of density $m(H)$ in H_{i+1} , must have also been in such a graph in H_i .

Since we are considering H such that, at least one vertex v , satisfies $m(v, H) < m(H)$, the above shows that the collapsing process will never produce a subgraph of density $m(H)$ containing these vertices and hence they will never be collapsed. This ensures that once all subgraphs have been collapsed, we will not be left with a single point, and that $m(\mathcal{H}) < m(H)$.

The above also demonstrates that while the choice of dense subgraph to collapse will result in different H_i , ultimately, this process will always terminate with the same final multigraph, which we call \mathcal{H} . To see why this follows, suppose a vertex lies in two different subgraphs, which we could choose to collapse. By (1) applied to the subgraph induced by the union of the two dense subgraphs, the new subgraph, formed by the collapsing process, will still have density $m(H)$ and so the remaining vertices will be collapsed at a later stage to the same point. \square

It is clear that if we can embed the collapsed, dense subgraphs of H , required for a factor, and then embed the edges of \mathcal{H} we will have our required factor. Firstly, we prove that we can embed these dense subgraphs as required. Consider the graph H' with vertex set $V(H)$ and edge set $E(H) - E(\mathcal{H})$, (i.e. H' contains only those edges collapsed by the above process). Let $\text{th}_{H'}(n)$ be the threshold for embedding a factor of H' into $G(n, p)$.

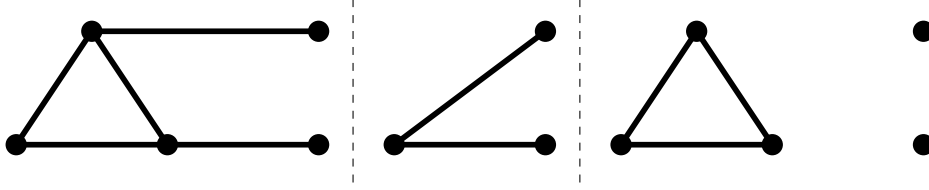


Figure 1: An example of an H and its respective \mathcal{H} and H' graphs

Lemma 3.2. *For a non-vertex-balanced graph H , and the corresponding H' as defined above, the threshold for the existence of an H' -factor satisfies;*

$$\text{th}_{H'}(n) = \Theta(n^{-1/m(H)}).$$

Proof. H' is a subgraph of H , and so $m(H') \leq m(H)$, and since H' contains copies of the collapsed subgraphs of density $m(H)$, it must itself satisfy $m(H') = m(H)$. The edges of H' are exactly those that were collapsed in the process that generated \mathcal{H} , and hence, any vertices of H that were not collapsed, will be isolated in H' . In other words, these vertices will have degree 0 in H' . Such vertices must exist, since we assumed that $m(v, H) < m(H)$ for some vertex v . Since we then have $\delta(H') < m(H')$, we can apply Theorem 2.2, completing the proof as required. \square

We now have an H -factor, without the edges from each copy of H , that are also present in \mathcal{H} . To embed these final edges, we now use our generalisation of 2.4, namely a specific application of Theorem 1.2 to find a factor of \mathcal{H} , with the collapsed subgraphs covered by the vertices we require from H .

As in [6], we work in (a multigraph generalised form of) $G(n, M)$, the graph chosen uniformly from all M -edge graphs on V (although we will use a multigraph generalised form of $G(n, p)$ to prove our results) and derive a generalised form of 2.3. Since we will be operating with multigraphs and partitioned vertex sets, we need to define some notation.

Let G be a graph on n vertices and let H be a fixed graph with vertex set $\{x_1, x_2, x_3, \dots, x_{v_H}\}$. Let \mathcal{H} be the multigraph obtained by repeated applications of vertex collapsing of subgraphs of H of density $m(H)$, until no such subgraphs remain. Let $k_{\mathcal{H}} = |V(\mathcal{H})|$ and $h_{\mathcal{H}} = |E(\mathcal{H})|$. If \mathcal{H} has vertex set $\{y_1, y_2, \dots, y_{k_{\mathcal{H}}}\}$, we use lemma 3.2 to find partial factors consisting of n/v_H copies of these collapsed subgraphs (or single vertices, for those that were not collapsed) for each vertex in \mathcal{H} .

We use the standard method of partitioning the edges of $G(n, p)$ into $G(n, p') \cup G(n, p')$ where there is an edge in $G(n, p)$ if and only if there is an

edge in at least one of the $G(n, p')$. Since we are only interested in threshold functions, which are equivalent up to constant factors, and $1 - p = (1 - p')^2$, which implies that $p' > p/2$, we can apply Lemma 3.2 without sacrificing randomness of the edges between these embedded subgraphs. This leaves us with $k_{\mathcal{H}}$ separate classes of the vertices of $G(n, p)$, each containing graphs of density $m(H)$ or isolated vertices, each corresponding to a vertex of \mathcal{H} .

We now wish to find an \mathcal{H} -factor between these partitioned sets, but we are only interested in factors that connect the ‘correct’ vertices together from each partition, and hence are not interested in the edges within each partition, or those that are not the prescribed edges between the subgraphs we have already embedded.

We can consider a random multigraph, which we call $\mathcal{H}(n, p)$, using the edges of our second $G(n, p')$, such that if the required \mathcal{H} -factor exists here, it will translate into the required H -factor in $G(n, p)$. Firstly, the vertex set of $\mathcal{H}(n, p)$ consists of a single vertex for each of the isolated vertices and subgraphs of density $m(H)$, that we have embedded into $G(n, p)$. We maintain the partition of these new vertices into equal sets of size n/v_H , according to the vertex of \mathcal{H} , they correspond to in the initial embedding. Note that this means that $\mathcal{H}(n, p)$ does not have n vertices, but rather $k_{\mathcal{H}}n/v_H$, which is however, a constant multiple of n .

For the edge set of $\mathcal{H}(n, p)$, we use the second edge partition $G(n, p')$, to ensure randomness. We consider a pair of vertices, v_1 and v_2 in $\mathcal{H}(n, p)$, noting that we can also consider v_1 and v_2 as sets of vertices of $G(n, p)$, and the mapping $\sigma : v(H) \rightarrow G(n, p)$ that describes the already embedded subgraphs that form the vertices of $\mathcal{H}(n, p)$. For each x_1 and $x_2 \in v(H)$, with $(x_1, x_2) \in e(H)$ and $\sigma(x_1) \in v_1$ and $\sigma(x_2) \in v_2$, if $(\sigma(x_1), \sigma(x_2)) \in e(G(n, p))$, then we have an edge between v_1 and v_2 in $\mathcal{H}(n, p)$, noting that we consider each such edge separately. In this way, any factor of \mathcal{H} found in $\mathcal{H}(n, p)$ will automatically translate into a factor of H in $G(n, p)$.

$\mathcal{H}(n, p)$ can also be thought of as a random $k_{\mathcal{H}}$ -partite multigraph with $k_{\mathcal{H}}n/v_H$ vertices, and edges between vertices x and y with probability p for each edge between their origin vertices in H , and 0 otherwise. Essentially, a series of $h_{\mathcal{H}} = e(\mathcal{H})$ bipartite graphs, in the same ‘shape’ as \mathcal{H} . We will work using this random graph model (and the corresponding $\mathcal{H}(n, p)$ model), to prove our results.

It may be helpful for some readers to visualise this as a $k_{\mathcal{H}}$ -partite random graph with different edge probabilities for some of the edge partitions, rather than a multigraph. For example, a single edge with probability p^2 , rather than two edges, each with probability p between vertices. The varying probabilities make this model cumbersome to work with, however, and the multigraph notation is more convenient for use.

For some choices of H it may simply be possible to set all edge probabilities to the minimum of these values, and still find the factor, but our earlier graph, containing a K_5 , is an example of a graph for which this method would fail.

At this point, we have the exact set-up for Theorem 1.2. As we have shown that $m(\mathcal{H}) < m(H)$, it implies that for $p = \omega(n^{-1/m(H)}) > \mathcal{O}(n^{-1/m(\mathcal{H})+o(1)})$ (or, if \mathcal{H} is strictly balanced $p > \omega(n^{-1/m(\mathcal{H})}(\log n)^{1/|E(\mathcal{H})|})$) such an \mathcal{H} -factor, a.a.s exists, and hence our H factor exists in our original $G(n, p)$, as required.

4 Theorem 1.2

The proof of Theorem 1.2 largely follows the same steps as the original graph result in [6]. To illustrate the key ideas, we outline the initial setup and then draw out several of the key ideas of the proof. We begin the proof of Theorem 1.2, with a multigraph generalised version of their Theorem 3.1. This Theorem essentially shows that the number of factors in $\mathcal{H}(n, p)$ is close to expectation, by demonstrating that the equivalent process of removing edges from the complete graph, does not remove too many factors at each step.

Theorem 4.1. *Let $k_{\mathcal{H}} = v(\mathcal{H})$ and $h_{\mathcal{H}} = e(\mathcal{H})$. For $p = p(n) = \omega(n^{-1/m(H)})$ and $M = M(n) = h_{\mathcal{H}}(n/v_H)^2 p$, and let $\Phi(G)$ be the number of the \mathcal{H} -Factors in a graph G , then*

$$\Pr(\Phi(\mathcal{H}(n, p)) \geq (n^{k_{\mathcal{H}}-1} p^{h_{\mathcal{H}}})^{n/v_H} e^{-\mathcal{O}(n)}) \geq 1 - n^{-\omega(1)}.$$

Proof. The proof follows in the same steps as the original. Let $T = h_{\mathcal{H}}(n/v_H)^2 - M$, and let $e_1, e_2, \dots, e_{h_{\mathcal{H}}(n/v_H)^2}$ be a random, uniform ordering of the complete form of our multigraph (i.e $\mathcal{H}(n, p)$ with $p = 1$), which we shall call KM_n . Set G_i to be $KM_n - \{e_1, e_2, \dots, e_i\}$, so if we let $\mathcal{F}(G)$ be the set of \mathcal{H} factors on G , we define $\mathcal{F}_i := \mathcal{F}(G_i)$. We then let ξ_i be the fraction of members of \mathcal{F}_i containing e_i . Then, as for the standard graph case, we have

$$|\mathcal{F}_t| = |\mathcal{F}_0| \frac{|\mathcal{F}_1|}{|\mathcal{F}_0|} \dots \frac{|\mathcal{F}_t|}{|\mathcal{F}_{t-1}|} = |\mathcal{F}_0| (1 - \xi_1) \dots (1 - \xi_t),$$

and that therefore, we have

$$\log |\mathcal{F}_t| = \log |\mathcal{F}_0| + \sum_{i=1}^t \log(1 - \xi_i). \quad (2)$$

Here our sums start to differ somewhat from the standard graph case; we have

$$\log |\mathcal{F}_0| = \log((n/v_H)!)^{k_{\mathcal{H}}-1} = \frac{k_{\mathcal{H}}-1}{v_H} n \log n - \mathcal{O}(n).$$

Also, we have

$$\mathbb{E} \xi_i = \frac{h_{\mathcal{H}} n / v_H}{h_{\mathcal{H}}(n/v_H)^2 - i + 1} =: \gamma_i = \mathbb{E}[\xi_i | e_1, \dots, e_{i-1}]$$

for any choice of e_1, \dots, e_{i-1} .

Therefore we have

$$\sum_{i=1}^t \mathbb{E} \xi_i = \sum_{i=1}^t \gamma_i = \frac{h_{\mathcal{H}} n}{v_H} \log \frac{h_{\mathcal{H}}(n/v_H)^2}{h_{\mathcal{H}}(n/v_H)^2 - t} + o(1)$$

provided that $h_{\mathcal{H}}(n/v_H)^2 - t > \omega(n)$. We use the same property as for the standard graph case, namely \mathcal{A}_t , which is the event that

$$\left\{ \log |\mathcal{F}_t| > \log |\mathcal{F}_0| - \sum_{i=1}^t \gamma_i - \mathcal{O}(n) \right\}.$$

As before, we aim to show that with high probability \mathcal{A}_t does not fail, i.e.

$$\text{for } t \leq T, \Pr(\overline{\mathcal{A}_t}) = n^{-\omega(1)}. \quad (3)$$

This implies our theorem, since we then, setting $t = T$, have

$$\log \Phi(\mathcal{H}(n, p)) = \log |\mathcal{F}_t| > \frac{k_{\mathcal{H}}-1}{v_H} n \log n + \frac{h_{\mathcal{H}} n}{v_H} \log p - \mathcal{O}(n)$$

(since $M = h_{\mathcal{H}}(n/v_H)^2 - T = h_{\mathcal{H}}(n/v_H)^2 p$). To prove (3), we use the same methods as [6], namely an Azuma's inequality, martingale argument. As before, we will define two auxiliary properties \mathcal{B}_i and \mathcal{R}_i for $i \leq 1 \leq T-1$, that will allow us to establish control over the concentration of our variables. We set our martingale to have a difference sequence of

$$Z_i = \begin{cases} \xi_i - \gamma_i & \text{if } \mathcal{B}_i \text{ and } \mathcal{R}_i \text{ hold for all } j < i \\ 0 & \text{otherwise.} \end{cases}$$

And so our martingale is $X_t = \sum_{i=1}^t Z_i$. We leave the formal definitions of \mathcal{B}_i and \mathcal{R}_i for Section 5.3, but in general terms, \mathcal{R}_i states that each vertex is in close to expectation, number of copies of \mathcal{H} , along with a second technical condition, while \mathcal{B}_i states that the maximum number of factors using

a particular copy of \mathcal{H} , is close to the average over all copies of \mathcal{H} . For all $i \leq t$, we will have that \mathcal{B}_{i-1} and \mathcal{R}_{i-1} imply

$$\xi_i = o(\log^{-1} n). \quad (4)$$

Our martingale analysis, will give us that $\Pr(|X_t| > n) < n^{-\omega(1)}$ (i.e. $|X_t| \leq \mathcal{O}(n)$), and if we have \mathcal{B}_i and \mathcal{R}_i for $i < t \leq T$, we will then have that $X_t = \sum_{i=1}^t \xi_i - \gamma_i$ and therefore

$$\sum_{i=1}^t \xi_i < \sum_{i=1}^t \gamma_i + \mathcal{O}(n) < \mathcal{O}(n \log n).$$

Using this, (2), (4) and the series expansion for $\log(1 - x)$ we get that

$$\log |\mathcal{F}_t| > \log |\mathcal{F}_0| - \sum_{i=1}^t (\xi_i + \xi_i^2) > \log |\mathcal{F}_0| - \sum_{i=1}^t \gamma_i - \mathcal{O}(n).$$

As in the graph case, we are left with three possibilities for the failure of this to occur and hence,

$$\Pr(\overline{\mathcal{A}}_t) < \sum_{i < t} \Pr(\overline{\mathcal{R}}_i) + \sum_{i \leq t} \Pr(\wedge_{j < i} (\mathcal{B}_j \mathcal{R}_j) \wedge \overline{\mathcal{A}}_i) + \sum_{i < t} \Pr(\mathcal{A}_i \mathcal{R}_i \overline{\mathcal{B}}_i).$$

The previously mentioned martingale analysis shows that the second term is at most $n^{-\omega(1)}$, and we follow the same processes as [6] in Section 5 to show that for $i \leq T$

$$\Pr(\overline{\mathcal{R}}_i) < n^{-\omega(1)} \quad (5)$$

and

$$\Pr(\mathcal{A}_i \mathcal{R}_i \overline{\mathcal{B}}_i) < n^{-\omega(1)}. \quad (6)$$

These three bounds give us the required result. \square

In the next sections we outline the generalisation of the various results from the original factors paper, and include some notes on how we can apply them to our multigraph situation. Rather than just present a slightly modified reproduction of [6], and to make this generalisation of more value to the reader, we have first drawn out what Johansson, Kahn and Vu referred to as ‘the heart of the matter’ and presented it as a stand-alone result, with our required generalisation and then continuing with the surrounding proofs.

As demonstrated above, the factor result follows from showing that

$$\Pr(\overline{\mathcal{R}}_i) < n^{-\omega(1)}$$

and

$$\Pr(\mathcal{A}_i \mathcal{R}_i \overline{\mathcal{B}}_i) < n^{-\omega(1)}.$$

In proving the second equation, a second graph property \mathcal{C} is introduced. The following shows that the failure of \mathcal{C} results in two sets differing by a single vertex, such that the number of factors on the complement of these sets (subject to some restrictions) vary significantly. The proof of this revolves around the use of entropy results that we describe below, while in Section 5.6, concentration results are used to demonstrate that the event is unlikely as required.

4.1 Entropy

We follow the results of chapter 6 of the factors paper but are left with a modification to make to their Lemma 6.1. As in the original, we have $H(X)$ to be the base entropy of a discrete random variable X , i.e.,

$$H(X) = \sum_x p(x) \log \frac{1}{p(x)},$$

where $p(x) = \Pr(x = X)$. Now, in our case, given a vertex y in a random multigraph G , we use $X(y, G)$ to be the copy of \mathcal{H} in a randomly chosen \mathcal{H} -factor, and that $h(y, G) = H(X(y, G))$. We will require a slightly different result, than in the original, as we will only be interested in vertices y from a single partition set of our random multigraph. We suppose that our multigraph has the same structure as $\mathcal{H}(n, p)$, i.e. any copy of \mathcal{H} will contain one vertex from each partition set of G . Given V_1 , a partition set of G , we have the following

Lemma 4.2.

$$\log \Phi(G) \leq \sum_{y \in V_1} h(y, G).$$

Proof. This result follows in the same way as in the original, using a variant statement of Shearer's Lemma. This states that, given a random vector $Y = (Y_i : i \in I)$, and \mathcal{S} , a collection of subsets of I , with repeats allowed, such that each element of I belongs to at least t members of \mathcal{S} . For $S \in \mathcal{S}$ let Y_S be the random vector $Y_i : i \in S$. Then $H(Y) \leq t^{-1} \sum_{S \in \mathcal{S}} H(Y_S)$. If we let Y be the indicator for the random \mathcal{H} -Factor, then I is the set of copies of \mathcal{H} in the complete form of our multigraph and \mathcal{S} is the collection of sets S_x , where S_x is the collection of copies of \mathcal{H} containing a vertex x , taken only

over $x \in V_1$. We have, therefore, that each copy of \mathcal{H} , belongs to exactly one S_x in \mathcal{S} , and so $t = 1$. It follows that;

$$H(Y) = \sum_{\Phi(G)} \frac{1}{\Phi(G)} \log(\Phi(G)) = \log(\Phi(G))$$

and since $H(Y_S) = h(y, G)$, the proof is complete. \square

The second entropy result of [6], namely Lemma 6.2, is not specific to graphs, and hence requires no generalisation for our uses. We state it below for reference.

We let S be a finite set, $W : S \rightarrow \mathcal{R}^+$, and let X be the random variable taking values in S with probability

$$\Pr(X = x) = W(x)/W(S),$$

where, for a set $A \subseteq S$, $W(A)$ is the sum of W over the members of A , i.e. $W(A) = \sum_{x \in A} W(x)$.

Lemma 4.3. *If $H(X) > \log |S| - \mathcal{O}(1)$, then there are $a, b \in \text{range}(W)$ with*

$$a \leq b < \mathcal{O}(a)$$

such that for $J = W^{-1}[a, b]$ we have,

$$|J| = \Omega(|S|)$$

and

$$W(J) > 0.7W(S).$$

4.2 The heart of the matter

We let $\Phi(G)$ be the number of \mathcal{H} -factors on a partitioned multigraph G , \mathcal{V}_0 be the set of vertex sets of size $k_{\mathcal{H}}$ in $\mathcal{H}(n, p)$ with a vertex from each partition set, and $\mathcal{H}(x, G)$ be the set of copies of \mathcal{H} in G containing the vertex x , again with each vertex from a separate partition set. We define $D(x, G) = |\mathcal{H}(x, G)|$ to be the number of copies of \mathcal{H} in G containing a vertex x , while $D(p)$ is the expectation of $D(x, \mathcal{H}(n, p))$ in $\mathcal{H}(n, p)$ given a randomly chosen x .

For Z , a disjoint union of elements of \mathcal{V}_0 , we define $w : \mathcal{V}_0 \rightarrow \mathcal{R}^+$ as $w(Z) = \Phi(\mathcal{H}(n, p) \setminus Z)$, i.e. the number of partial \mathcal{H} -factors in $\mathcal{H}(n, p)$ with only the vertices in Z , not covered.

Lastly, fixing a set of vertices, Y of size $k_{\mathcal{H}} - 1$, taken from separate partition sets and two vertices x and y , both from the remaining partition

set in $V(\mathcal{H}(n, p) \setminus Y)$ we define $w_{x,y} : \mathcal{H}(x, \mathcal{H}(n, p) - \{Y \cup \{y\}\}) \rightarrow \mathcal{R}^+$ as $w_{x,y}(K) = w(K \cup Y \cup \{y\})$. In simple terms, $w_{x,y}$ can be thought of as the number of H -factors on $\mathcal{H}(n, p) \setminus \{Y \cup \{y\}\}$ that use K as the copy of \mathcal{H} containing x in the \mathcal{H} -factor.

Definition 1. We say $\mathcal{H}(n, p)$ satisfies $\mathcal{A}(p)$ if the following holds

$$\log(\Phi(\mathcal{H}(n, p))) > \frac{k_{\mathcal{H}} - 1}{v_H} n \log n + \frac{h_{\mathcal{H}} n}{v_H} \log p - \mathcal{O}(n).$$

Definition 2. We say that $\mathcal{H}(n, p)$ satisfies $\mathcal{R}_b(p)$ if the following condition holds.

For each $x \in V$, $|D(x, \mathcal{H}(n, p)) - D(p)| = o(D(p))$.

Informally, $\mathcal{A}(p)$ says that the number of factors is close to expectation, while $\mathcal{R}_b(p)$ says the same for the number of copies of \mathcal{H} that each vertex of $\mathcal{H}(n, p)$ is in.

For a $k_{\mathcal{H}} - 1$ subset of $V(\mathcal{H}(n, p))$, as always with each vertex taken from different partition sets; Y , let $\mathcal{V}_0(Y)$ be the set of $k_{\mathcal{H}}$ subsets containing Y , with the final vertex taken from the remaining partition set.

Definition 3. We define \mathcal{C} for $\mathcal{H}(n, p)$ as follows: $\mathcal{H}(n, p)$ satisfies \mathcal{C} if for all $k_{\mathcal{H}} - 1$ subsets of $\mathcal{H}(n, p)$, Y as above, we have the following:

$$\max w(\mathcal{V}_0(Y)) \leq \max\{n^{-2(k_{\mathcal{H}}-1)}\Phi(\mathcal{H}(n, p)), 2\text{med}w(\mathcal{V}_0(Y))\}$$

We prove the following Theorem;

Theorem 4.4. $\mathcal{AR}_b\overline{\mathcal{C}}$ implies that there exists a set of vertices Y , each taken from $|Y| = v_H - 1$ different partition sets of $\mathcal{H}(n, p)$, and x, y in the remaining partition set, such that we can find a collection J of elements of $H(y, \mathcal{H}(n, p) - (Y \cup \{x\}))$ and J' from $H(x, \mathcal{H}(n, p) - (Y \cup \{y\}))$ with $|J| > \Omega(|H(y, \mathcal{H}(n, p) - (Y \cup \{x\}))|)$, and $w_{y,x}^{-1}|J| = w_{x,y}^{-1}|J'| = [a, b]$ with $a \leq b < \mathcal{O}(a)$ satisfying

$$\sum_{X \in J} w_{y,x}(X) > 0.7w(Y \cup \{x\})$$

and

$$\sum_{X \in J'} w_{x,y}(X) \leq 0.5w(Y \cup \{x\})$$

Proof. Suppose that \mathcal{A} , \mathcal{R}_b hold but that \mathcal{C} fails. Therefore we can find at least one set Y at which \mathcal{C} fails. We therefore know that there exists x , such that $w(Y \cup \{x\})$ is maximum for choices of x and satisfies

$$w(Y \cup \{x\}) > n^{-2(k_{\mathcal{H}}-1)}\Phi(\mathcal{H}(n, p)).$$

We now choose y with $w(Y \cup \{y\}) \leq \text{med } w(\mathcal{V}_0(Y))$, and $h(y, \mathcal{H}(n, p) - (Y \cup \{x\}))$ maximal, given this constraint.

Given \mathcal{A} , we know that

$$\log(\Phi(\mathcal{H}(n, p))) > \frac{k_{\mathcal{H}} - 1}{v_H} n \log n + \frac{h_{\mathcal{H}} n}{v_H} \log p - \mathcal{O}(n).$$

While the failure of \mathcal{C} tells us that

$$w(Y \cup \{x\}) = \Phi(\mathcal{H}(n, p) - (Y \cup \{x\})) > n^{-2(k_{\mathcal{H}}-1)} \Phi(\mathcal{H}(n, p)),$$

hence, combining the two we have,

$$\log \Phi(\mathcal{H}(n, p) - (Y \cup \{x\})) > \frac{k_{\mathcal{H}} - 1}{v_H} n \log n + \frac{h_{\mathcal{H}} n}{v_H} \log p - \mathcal{O}(n). \quad (7)$$

We use Lemma 4.2 and apply it to the graph $\mathcal{H}(n, p) - (Y \cup \{x\})$. Letting V_1 be the partition set containing x and y , we have,

$$\log \Phi(\mathcal{H}(n, p) - (Y \cup \{x\})) \leq \sum_{z \in V_1} h(z, \mathcal{H}(n, p) - (Y \cup \{x\})). \quad (8)$$

However, we know that we chose y to have maximal entropy, chosen from a set of at least half of possible such z , and we also have that for any random variable X , the entropy $H(X) \leq \log(|\text{range}(X)|)$ (with equality only if the variable is uniformly distributed).

The range of our random variable is contained in the set of copies of H containing the fixed vertex z in $\mathcal{H}(n, p) - (Y \cup \{x\})$, which by definition, is of size $D(z, \mathcal{H}(n, p) - (Y \cup \{x\})) \leq D(z, \mathcal{H}(n, p))$. We know from \mathcal{R}_b that this is less than $(1 + o(1))D(p)$. Hence, we have that at least half the z 's in (8) satisfy $h(z, \mathcal{H}(n, p) - (Y \cup \{x\})) \leq h(y, \mathcal{H}(n, p) - (Y \cup \{x\}))$ and the remaining, $n/2v_H$ all satisfy $h(z, \mathcal{H}(n, p) - (Y \cup \{x\})) \leq \log((1 + o(1))D(p))$. Therefore, (8) gives us,

$$\begin{aligned} \log(\Phi(\mathcal{H}(n, p) - (Y \cup \{x\}))) &\leq \sum_{z \in V_1} h(z, \mathcal{H}(n, p) - (Y \cup \{x\})) \\ &\leq \frac{n}{2v_H} (h(y, \mathcal{H}(n, p) - (Y \cup \{x\})) + \log(1 + o(1))D(p)) \\ &\leq \frac{n}{2v_H} (h(y, \mathcal{H}(n, p) - (Y \cup \{x\})) + (k_{\mathcal{H}} - 1) \log n + h_{\mathcal{H}} \log p). \end{aligned}$$

Rearranging, to get $h(y, \mathcal{H}(n, p) - (Y \cup \{x\}))$ on the left, and substituting from (7) we have,

$$h(y, \mathcal{H}(n, p) - (Y \cup \{x\})) \geq (k_{\mathcal{H}} - 1) \log n + h_{\mathcal{H}} \log p - \mathcal{O}(1).$$

By \mathcal{R}_b we have that

$$\begin{aligned} \log(D(y, \mathcal{H}(n, p) - (Y \cup \{x\}))) &\leq \log D(y, \mathcal{H}(n, p)) \leq \log((1 + o(1))D(p)) \\ &= (k_{\mathcal{H}} - 1) \log n + h_{\mathcal{H}} \log p + \log(1 + o(1)), \end{aligned}$$

and hence combining with the above, we have

$$h(y, \mathcal{H}(n, p) - (Y \cup \{x\})) > \log(D(y, \mathcal{H}(n, p) - (Y \cup \{x\}))) - \mathcal{O}(1). \quad (9)$$

We now use our functions $w_{y,x}$ and $w_{x,y}$, previously defined as;

$$w_{y,x}(K) = w(K \cup Y \cup \{x\}) \text{ and similarly } w_{x,y}(K) = w(K \cup Y \cup \{y\}).$$

With $w_{y,x}$ defined on $H(y, \mathcal{H}(n, p) - (Y \cup \{x\}))$; the set of copies of \mathcal{H} containing y in $\mathcal{H}(n, p) - (Y \cup \{x\})$, and similarly, $w_{x,y}$ defined on $H(x, \mathcal{H}(n, p) - (Y \cup \{y\}))$. Simply put, for a copy of \mathcal{H} containing y in $\mathcal{H}(n, p) - (Y \cup \{x\})$, $w_{y,x}$ is the number of \mathcal{H} -factors on this set, using that copy of \mathcal{H} .

If we consider the random variable $X(y, \mathcal{H}(n, p) - (Y \cup \{x\}))$, which is the copy of \mathcal{H} containing y in a uniformly at random chosen \mathcal{H} -factor on $\mathcal{H}(n, p) - (Y \cup \{x\})$, we can see that the probability that $X(y, \mathcal{H}(n, p) - (Y \cup \{x\})) = \mathcal{H}'$ for $\mathcal{H}' \in \mathcal{H}(y, \mathcal{H}(n, p) - (Y \cup \{x\}))$, is

$$w_{y,x}(\mathcal{H}') / \sum_{Z \in \mathcal{H}(y, \mathcal{H}(n, p) - (Y \cup \{x\}))} w_{y,x}(Z).$$

Also note that the denominator is equal to $w(\mathcal{H}(n, p) - (Y \cup \{x\}))$, since by summing only over copies of \mathcal{H} , we are counting each \mathcal{H} -factor exactly once.

Similarly, $X(x, \mathcal{H}(n, p) - (Y \cup \{y\}))$ is determined by $w_{x,y}$, and the sum $\sum_Z w_{x,y}(Z)$ is equal to $w((Y \cup \{y\}))$.

By the above, we have the setup used for Lemma 4.3, with $S = \mathcal{H}(y, \mathcal{H}(n, p) - (Y \cup \{x\}))$. Noting that $|S| = D(y, \mathcal{H}(n, p) - (Y \cup \{x\}))$, (9) gives us the required condition, and we are able to apply the result to $w_{y,x}$. This implies that there exist a and $b \in \text{range}(w_{y,x})$, for which we can set $J := w_{y,x}^{-1}([a, b])$, and it will satisfy the following:

$$|J| > \Omega(|\mathcal{H}(y, \mathcal{H}(n, p) - (Y \cup \{x\}))|)$$

and

$$\sum_{z \in J} w_{y,x}(Z) > 0.7 \sum_{z \in H(y, G - (Y \cup \{x\}))} w_{y,x}(Z) = 0.7 w(Y \cup \{x\}).$$

In simple terms, J is of the same magnitude in size as the whole pre-image of $w_{y,x}$, and its elements have overall weight at least a constant multiple of that of the whole set.

Equally we can set $J' = w_{x,y}^{-1}([a, b])$, and we know that

$$\sum_{J'} w_{x,y}(Z) \leq w(Y \cup \{y\}) < 0.5w(Y \cup \{x\})$$

The first inequality follows from simply summing over the full set containing J' , and the 2nd from our original definition of y and x . This completes the proof. \square

Proving that this is a.a.s. unlikely to happen, requires a range of concentration and technical lemmas, demonstrated in the following sections. In applying this result to Shamir's problem, if instead of considering factors, a matching of hyperedges is required, it has been shown that the proof follows with much more ease using a union bound argument, reducing the technical complexity of the proof considerably [4].

5 Generalisation of remaining results from [6]

The generalisation to partitioned structures and multigraphs of the remaining results and properties of [6], follow largely from careful consideration of sums and bounds, and formulation of polynomials. The following sections follow the structure of [6] closely, and are largely a technical exercise, that offer little to those who have read the original paper.

To highlight why the generalisation should follow, we note that while limiting the factors to these partitions appears to drastically limit the number of possible copies of H , since each partition is of linear in n size, we still have $\mathcal{O}(n^{k_{\mathcal{H}}}) = \mathcal{O}\binom{n}{k_{\mathcal{H}}}$, possible choices of vertices for each H , as in the standard case.

We also address the threshold required for applying Theorem 1.2 in obtaining Theorem 1.1. We are not guaranteed strict balance for the resulting \mathcal{H} , but regardless, the collapsing process, eliminates all subgraphs of density $m(H)$, and hence, $m(\mathcal{H}) < m(H)$ and so for $p > n^{-1/m(H)}$, we have a greater probability than required within the proof and so with Theorem 4.1, applied to $\mathcal{H}(n, p)$, on the partial factors already embedded during the collapsing process, we have Theorem 1.1 as required.

Throughout the proofs, for clarity in understanding our main result, we treat \mathcal{H} , as the graph formed by the collapsing process on some H , and that we have $p > n^{-1/m(H)}$ as in Theorem 1.1, but for proving Theorem 1.2 in full generality, \mathcal{H} may not necessarily be derived from some H , and we only have $p > n^{-1/m(\mathcal{H})+o(1)}$ (or with a log term for the strictly balanced case).

In this case the proof is unchanged, as throughout, as in [6], we only require that if $p = \mathcal{O}(p^{-1/a})$, \mathcal{H} contains no subgraphs of density equal to $1/a$, and that $n^{k_{\mathcal{H}}-1}p^{h_{\mathcal{H}}} = \omega(\log n)$, which follows immediately from the conditions in Theorem 1.2, given that the $o(1)$ term decreases sufficiently slowly.

5.1 Concentration Results

Firstly we address the usage of the various concentration results in Section 5 of [6]. These results are largely special cases of results by V. Vu that can be found in [7] (with J.H.Kim), [14] and [15].

We will utilise the various polynomial results here without modification, and hence will not repeat the proofs here again.

We will require some of the notation used in this section later, which we outline now. We let $f = f(t_1, t_2, \dots, t_n)$, be a polynomial of degree d with real coefficients. We say f is normal if its coefficients are positive, with the maximum coefficient being 1, and we note that the results here are also true for $\mathcal{O}(1)$ normal polynomials, which simply means that the polynomial's coefficients have some fixed bound.

We will consider polynomials that are multilinear which means we can express f in the form $f(t) = \sum \alpha_U t_U$, where U ranges over subsets of $[n]$ and $t_U := \prod_{u \in U} t_u$.

Lastly, we need that for a set $L \subseteq [n]$, the partial derivative of order $|L|$ with respect to the variables indexed by L is $\sum_{U \supseteq L} \alpha_U t_{U \setminus L}$, and its expectation, denoted \mathbb{E}_L or $\mathbb{E}_L f$ is $\sum \{\alpha_U \prod_{i \in U \setminus L} p_i : U \supseteq L\}$, where $t_i \sim \text{Ber}(p_i)$. Set $\mathbb{E}_j f = \max_{|L|=j} \mathbb{E}_L f$. We write $\mathbb{E}'_L = \mathbb{E}'_L f$ for the expectation of the non-constant part of the partial derivative of f , with respect to L , noting that for homogeneous, f of degree d , and $0 < |L| < d$, we have $\mathbb{E}'_L f = \mathbb{E}_L f$.

We take the original example used to illustrate the usage of these results, namely that we consider our polynomial f to be the number of copies of \mathcal{H} in our random multigraph $\mathcal{H}(n, p)$ containing a particular, fixed vertex x_0 .

We have that $f = \sum_U t_U$ where U runs over edge sets of copies of \mathcal{H} in our complete multigraph, containing our vertex x_0 . We have $\mathbb{E} f = \Theta(n^{k_{\mathcal{H}}-1}p^{h_{\mathcal{H}}})$, while for any non empty subset L of edges from the complete graph, the partial derivative will be $\mathbb{E}_L f = \sum_{U \supseteq L} t_{U \setminus L}$. This will be 0 for L that do not satisfy our multigraph structure requirements (i.e. at most one edge from each bipartite pairing forming the multigraph), in the same way that choosing an L not forming a subgraph of H would do, in the standard graph case.

In all theorems in the chapter we are interested in ensuring that the maximum value of the derivative does not exceed a certain magnitude, and

so in this sense, we are not interested in these cases, and so they cause no issue in this generalisation.

Given that our choice of L does satisfy our structure requirements, (and hence will be contained within at least one copy of \mathcal{H} in the complete graph), we can consider the graph formed by the edges of L , and the vertex end-points of these edges. Letting $k'_{\mathcal{H}}$ and $h'_{\mathcal{H}}$ be the number of vertices and edges respectively of L , then if L contains x_0 , we have $\mathbb{E}_L f = \mathcal{O}(n^{k_{\mathcal{H}}-k'_{\mathcal{H}}} p^{h_{\mathcal{H}}-h'_{\mathcal{H}}})$, and $\mathcal{O}(n^{k_{\mathcal{H}}-k'_{\mathcal{H}}-1} p^{h_{\mathcal{H}}-h'_{\mathcal{H}}})$ otherwise. Either way, we have,

$$\mathbb{E}f/\mathbb{E}_L f = \Omega(n^{k'_{\mathcal{H}}-1} p^{h'_{\mathcal{H}}}).$$

While we do not have strict balance of \mathcal{H} , we do have that it contains no subgraphs of density $m(H)$, and hence we have that $h'_{\mathcal{H}}/(k'_{\mathcal{H}}-1) < m(H)$ and recalling that $p = \omega(n^{-1/m(H)})$, we have that $\mathbb{E}f = \Omega(1)$ and that $\mathbb{E}f/\mathbb{E}_L f \geq n^{\Omega(1)}$, as is required for applying the results in the chapter.

For reference we include the results from the chapter below.

Theorem 5.1. *The following holds for any fixed positive integer d and positive constant ϵ . Let f be a multilinear, homogeneous, normal polynomial of degree d such that $\mathbb{E}f \geq n^\epsilon \max_{1 \leq j \leq d} \mathbb{E}_j f$. Then*

$$\Pr(|f - \mathbb{E}f| > \epsilon \mathbb{E}f) = n^{-\omega(1)}.$$

Theorem 5.2. *The following holds for any fixed positive integer d and positive constant ϵ . Let f be a multilinear, normal, homogeneous polynomial of degree d such that $\mathbb{E}f = \omega(\log n)$ and $\max_{1 \leq j \leq d-1} \mathbb{E}_j f \leq n^\epsilon$. Then*

$$\Pr(|f - \mathbb{E}f| > \epsilon \mathbb{E}f) = n^{-\omega(1)}.$$

Theorem 5.3. *The following holds for any fixed positive integer d and positive constant ϵ . Let f be a multilinear, normal, homogeneous polynomial of degree d such that $\mathbb{E}f = \omega(\log n)$ and $\max_{1 \leq j \leq d-1} \mathbb{E}_j f \leq n^\epsilon \mathbb{E}f$. Then*

$$\Pr(|f - \mathbb{E}f| > \epsilon \mathbb{E}f) = n^{-\omega(1)}.$$

Corollary 5.4. *The following holds for any fixed positive integer d and positive constant ϵ . Let f be a multilinear, normal, homogeneous polynomial of degree d such that $\mathbb{E}f \leq A$ where $A = A(n)$ satisfies*

$$A \geq \omega(\log n) + n^\epsilon \max_{0 < j < d} \mathbb{E}_j f,$$

then

$$\Pr(f > (1 + \epsilon)A) \leq n^{-\omega(1)}.$$

Theorem 5.5. *The following holds for any fixed positive integer d and positive constant ϵ . Let f be a multilinear, normal polynomial of degree d with $\mathbb{E}f = \omega(\log n)$ and $\max_{L \neq \emptyset} \mathbb{E}'_L f \leq n^\epsilon \mathbb{E}f$. Then*

$$\Pr(|f - \mathbb{E}f| > \epsilon \mathbb{E}f) \leq n^{-\omega(1)}.$$

Corollary 5.6. *The following holds for any fixed positive integer d and positive constant ϵ . Let f be a multilinear, normal polynomial of degree d such that $\mathbb{E}f \leq A$ where $A = A(n)$ satisfies*

$$A \geq \omega(\log n) + n^\epsilon \max_{L \neq \emptyset} \mathbb{E}'_L f,$$

then

$$\Pr(f > (1 + \epsilon)A) \leq n^{-\omega(1)}.$$

Theorem 5.7. *The following holds for any fixed positive integer d and positive constant ϵ . Let f be a multilinear, normal polynomial of degree d with $\max_L \mathbb{E}'_L f \leq n^\epsilon$. Then for any $\beta(n) = \omega(1)$,*

$$\Pr(f > \beta(n)) = n^{-\omega(1)}.$$

5.2 Martingale

The proof of the bound on the martingale follows exactly as that for the original paper. We have all the same bounds, namely that $|Z_i| < \varepsilon := \log^{-1} n$, and that $\sum_{i=1}^t \gamma_i = \mathcal{O}(n \log n)$, and the proof makes no use of the graph setting for the problem. We include the steps below for reference.

We let $X_t = Z_1 + \dots + Z_t$, and aim to show that

$$\Pr(X_t \geq n) < n^{-\omega(1)}.$$

We have that Z_i is a function of the random sequence e_1, \dots, e_i , but that $\mathbb{E}(Z_i | e_1, \dots, e_{i-1}) = 0$ for any choice of the e_j 's. Using (4), it will follow from the properties \mathcal{R} and \mathcal{B} , that $|Z_i| < \varepsilon := \log^{-1} n$. We can apply Markov's inequality to derive the following, for any positive h ;

$$\Pr(X_t \geq n) = \Pr(e^{h(Z_1 + \dots + Z_t)} \geq e^{hn}) \leq \mathbb{E}(e^{h(Z_1 + \dots + Z_t)})e^{-hn}. \quad (10)$$

Using $Z_i = \xi_i - \gamma_i$, we have that $\mathbb{E}(\xi_i | e_1, \dots, e_{i-1}) = \gamma_i$. Using $0 \leq \xi_i \leq \varepsilon$ and the convexity of e^x , we have

$$\mathbb{E}(e^{hZ_i} | e_1, \dots, e_{i-1}) \leq e^{-h\gamma_i} \left(\left(1 - \frac{\gamma_i}{\varepsilon}\right) + \frac{\gamma_i}{\varepsilon} e^{h\varepsilon} \right).$$

Taylor series expansions, show that the right hand side is at most $e^{h^2\epsilon\gamma_i}$, for any $0 \leq h \leq 1$. Using induction on t we derive the following.

$$\begin{aligned}\mathbb{E}(e^{h(Z_1+\dots+Z_t)}) &= \mathbb{E}(\mathbb{E}(e^{h(Z_1+\dots+Z_t)}|e_1, \dots, e_{t-1})) \\ &= \mathbb{E}(e^{h(Z_1+\dots+Z_{t-1})}\mathbb{E}(e^{hZ_t}|e_1, \dots, e_{t-1})) \\ &\leq \mathbb{E}(e^{h(Z_1+\dots+Z_{t-1})}e^{h^2\epsilon\gamma_t}) \\ &\leq e^{h^2\epsilon\sum_{i=1}^t\gamma_i}.\end{aligned}$$

Combined with (10), we have

$$\Pr(X_t \geq n) \leq e^{h^2\epsilon\sum_{i=1}^t\gamma_i - hn}.$$

We have that $\sum_{i=1}^t\gamma_i = \mathcal{O}(n \log n)$ and $\epsilon = \log^{-1} n$, and so setting h to be a sufficiently small positive constant, leaves the right hand side as $e^{-\Omega(n)} = n^{-\omega(1)}$, as required.

5.3 The Properties \mathcal{B} and \mathcal{R}

We now define our slightly altered properties \mathcal{B}_i and \mathcal{R}_i . We note, as before that in proving (5) and (6) we can operate in the random graph $\mathcal{H}(n, p_i)$ rather than G_i , where

$$p_i = 1 - \frac{i}{h_{\mathcal{H}}(n/v_H)^2}.$$

We will define graph properties \mathcal{B} and $\mathcal{R}(p)$ and then the event \mathcal{B}_i will be $\{G_i \text{ satisfies } \mathcal{B}\}$ and \mathcal{R}_i will be $\{G_i \text{ satisfies } \mathcal{R}(p_i)\}$.

In defining \mathcal{B} , we use the same notation for the functions W , namely that for a finite set A and $W : A \rightarrow [0, \infty)$, set

$$\overline{W}(A) = |A|^{-1} \sum_{a \in A} W(a),$$

$$\max W(A) = \max_{a \in A} W(a),$$

and lastly that,

$$\max_{\mathcal{R}} W(A) = \overline{W}(A)^{-1} \max W(A),$$

with $\text{med } W(A)$ the median of W on A .

For a multigraph G with our required partition structure, and vertex set V , let Z be a choice of $k_{\mathcal{H}}$ vertices from V , with each element taken from a different vertex partition. We then let $w_G(Z) = \Phi(G - Z)$. Therefore, $w_G(Z)$ is the number of \mathcal{H} -factors in the multigraph induced by G on the vertex set $V \setminus Z$.

It could also be thought of as the number of \mathcal{H} -factors in G , containing Z as a copy of \mathcal{H} , if all edges between the vertices of Z had been added in, where they are not already present.

We also use $w_G(K) = w_G(V(K))$ for $K \in \mathcal{H}(G)$, the set of copies of \mathcal{H} in G (which will contain one vertex from each partition of G).

We now define property \mathcal{B} , for a multigraph G as for the graph case, namely

$$\mathcal{B}(G) = \{\max w_G(\mathcal{H}(G)) = \mathcal{O}(1)\}. \quad (11)$$

As in the graph case, $\mathcal{B}(G)$ states that no copy of \mathcal{H} in G is contained in much more than the average number of \mathcal{H} -factors, for a copy of \mathcal{H} .

We define $\mathcal{R}(p)$, for the most part, in the same manner as for a graph, with two parts to the definition. For the first part, we use almost the same set-up. We have G , our random multigraph and V its vertex set, and given $A \subseteq V(\mathcal{H})$, $E' \subseteq E(\mathcal{H}) \setminus E(\mathcal{H}[A])$, an injection ψ , from A to V (mapping vertices to the correct partition of V corresponding to their location in \mathcal{H}). We let $X(G)$ be the number of injections $\phi : V(\mathcal{H}) \rightarrow V$ with

$$\phi \equiv \psi \text{ on } A \quad (12)$$

and

$$xy \in E' \Rightarrow \phi(x)\phi(y) \in E(G).$$

We can write $X(\mathcal{H}(n, p))$ in an obvious way, as a polynomial in variables $t_e = \mathbf{1}_{\{e \in E(\mathcal{H}(n, p))\}}$, for e , an edge in the complete form of our multigraph:

$$X(\mathcal{H}(n, p)) = q(t) = \sum_{\phi} t_{\phi(E')},$$

where t is the indicator function for edges of the multigraph, and the sum is over all injections ϕ satisfying (12).

As in the original paper, we have that this function $q(t)$, is multilinear, $\mathcal{O}(1)$ -normal and homogeneous of degree $d = |E'|$. We use the same definition for \mathbb{E}^* ;

$$\mathbb{E}^* = \max\{\mathbb{E}_L q : |L| < d\}. \quad (13)$$

We also use the same definition for $D(p)$, using it as the expected number of copies of \mathcal{H} in $\mathcal{H}(n, p)$ using a given vertex $x \in V$, while $D(x, G)$ is the actual number of copies containing x in G . It is clear that

$$D(p) = (n/v_H)^{k_{\mathcal{H}}-1} p^{h_{\mathcal{H}}} = \Theta(n^{k_{\mathcal{H}}-1} p^{h_{\mathcal{H}}}).$$

We are now ready to define $\mathcal{R}(p)$, which is identical to the graph formulation, not taking into account our slight changes to the above notation.

Definition 4. We say that a random multigraph G satisfies $\mathcal{R}(p)$ if the following two conditions hold.

- (a) For A , E' and ψ (and associated notation) as above: if $\mathbb{E}^* = n^{-\Omega(1)}$, then for any $\beta(n) = \omega(1)$, $X(G) < \beta(n)$ for large enough n ; if $E^* \geq n^{-o(1)}$, then for any fixed $\epsilon > 0$ and large enough n , $X(G) < n^\epsilon \mathbb{E}^*$.
- (b) For each $x \in V$, $|D(x, G) - D(p)| = o(D(p))$

We can now prove that these conditions give the required bound on the size of ξ_i . The proof follows in the same fashion as in the original paper:

Lemma 5.8. For $i \leq T = h_{\mathcal{H}}(n/v_H)^2 - M$ as defined in Section 3, \mathcal{B}_{i-1} and \mathcal{R}_{i-1} (i.e. that G_{i-1} satisfies \mathcal{B} and $\mathcal{R}(p_i)$) imply (4)

Proof. Write w for $w_{G_{i-1}}$. We aim to show that \mathcal{B}_{i-1} and \mathcal{R}_{i-1} imply that, for any $K \in \mathcal{H}(G_{i-1})$,

$$w(K)/\Phi(G_{i-1}) = \mathcal{O}(1/D(p_{i-1})).$$

As before, the left hand side of this equation is the fraction of \mathcal{H} -factors in G_{i-1} that use K , and we prove this result in the same fashion, since we have;

$$\begin{aligned} \Phi(G_{i-1}) &= \frac{v_{\mathcal{H}}}{n} w(\mathcal{H}(G_{i-1})) \\ &= \frac{v_{\mathcal{H}}}{n} \Omega(|\mathcal{H}(G_{i-1})| \max w(\mathcal{H}(G_{i-1}))) \\ &= \Omega(D(p_{i-1}) w(K)). \end{aligned}$$

The first line follows, since each \mathcal{H} -factor will be counted n/v_H times by summing the w function over all copies of \mathcal{H} . The second line follows from applying \mathcal{B} , while the third comes from part (b) of \mathcal{R}_{i-1} , and noting that $(k_{\mathcal{H}} n/v_H) D(p) = k_{\mathcal{H}} |\mathcal{H}(G_{i-1})|$.

We also have, using the same arguments as in the original, that part (a) of \mathcal{R}_{i-1} implies that the number of $K \in \mathcal{H}(G_{i-1})$, containing a given edge $e \in E(G_{i-1})$ is at most $\beta(n)$, satisfying $\beta^{-1} D(p_{i-1}) = \omega(\log n)$. Here we also use that $D(p_{i-1}) = n^{k_{\mathcal{H}}-1} p_{i-1}^{h_{\mathcal{H}}} = \omega(\log n)$, for $i \leq T$. \square

Lastly we state the required ' p -version' of \mathcal{A}_t :

$$\mathcal{A}(p) = \left\{ \log |\mathcal{F}(G)| > \log |\mathcal{F}_0| - \sum_{i=1}^t \gamma_i - \mathcal{O}(n) \right\},$$

where $t = \lceil (1-p) h_{\mathcal{H}}(n/v_H)^2 \rceil$. Recall from the end of Theorem 4.1, that our required result will follow from the following lemmas

Lemma 5.9. For $p > \omega(n^{-1/m(H)})$,

$$\Pr(\mathcal{H}(n, p) \text{ satisfies } \mathcal{R}(p)) = 1 - n^{-\omega(1)}.$$

Lemma 5.10. For $p > \omega(n^{-1/m(H)})$,

$$\Pr(\mathcal{H}(n, p) \text{ satisfies } \mathcal{A}(p)\mathcal{R}(p)\overline{\mathcal{B}}) = n^{-\omega(1)}.$$

We prove these lemmas in the next subsections.

5.4 Regularity

We now prove Lemma 5.9, i.e. that with probability $1 - n^{-\omega(1)}$, $G = \mathcal{H}(n, p)$ satisfies both parts of the definition of $\mathcal{R}(p)$.

Part (a) follows easily in the same fashion as the original paper. There are only $n^{\mathcal{O}(1)}$ choice for each of A , E' and ψ , so we simply show that the probability of one of these violating (a) is $n^{-\omega(1)}$. As we noted earlier, we can express the random variable as a polynomial, and since it is homogeneous, multilinear, and $\mathcal{O}(1)$ -normal, we can apply the probability results from the concentration chapter of [6] directly, which gives us the result as required.

Proving (b) requires only slightly more adaptation to our scenario. We again express $D(x, G)$ as a polynomial of degree $h_{\mathcal{H}} = |E(\mathcal{H})|$, in the variables $t_e = \mathbf{1}_{e \in \mathbf{E}(\mathbf{G})}$, where e belongs to the complete form of our random multigraph. Therefore we have,

$$D(x, G) = f(t) := \sum \{t_K : K \in \mathcal{H}_0(x)\},$$

where $\mathcal{H}_0(x)$ is the set of copies of \mathcal{H} , containing x , in the complete multigraph, and $t_K = \prod_{e \in K} t_e$.

As noted earlier, we have,

$$\mathbb{E}f = (n/v_H)^{k_{\mathcal{H}}-1}p^{h_{\mathcal{H}}} = \Theta(n^{k_{\mathcal{H}}-1}p^{h_{\mathcal{H}}}) = \omega(\log n).$$

We aim to use one of the concentration results from [6], namely Theorem 5.3, which requires the above and that $\max_{1 \leq j \leq d-1} \mathbb{E}_j f \leq n^{-\epsilon} \mathbb{E}f$. For L a subset of the edges of the complete multigraph, with $1 \leq |L| = l < h_{\mathcal{H}}$, we have

$$\mathbb{E}_L f = p^{h_{\mathcal{H}}-l} N(L),$$

where $N(L)$ is the number of $K \in \mathcal{H}_0(x)$ with $L \subseteq E(K)$. Let $I = V(L) \cup \{x\}$, $V(L)$ being the set of vertices incident to the edges of L , and $k'_{\mathcal{H}} = |I|$. Then $N(L) = \Theta(n^{k_{\mathcal{H}}-k'_{\mathcal{H}}})$ if the graph $\mathcal{H}' := (I, L)$ is isomorphic to a subgraph of

\mathcal{H} , and zero otherwise. Therefore we have, recalling that $p > n^{-1/m(H)}$ and $d(H) = e_H/(v_H - 1)$,

$$\begin{aligned} \mathbb{E}f/\mathbb{E}_L f &= \Omega(n^{k'_H-1}p^l) = \Omega(n^{[(k'_H-1)/l-1/m(H)]l}) \\ &= \Omega(n^{[1/d(H)-1/m(H)]l}) = n^{\Omega(1)}. \end{aligned}$$

Using the fact that \mathcal{H} contains no subgraphs of density $m(H)$ (or denser). We therefore have the required conditions to use the concentration theorem and the result follows, which provides us with part (b) of \mathcal{R} as required.

5.5 Proof of Lemma 5.10

We now begin the proof of Lemma 5.10, continuing in the same vein as the original paper, as there, we will prove that \mathcal{B} is satisfied, using an auxiliary event, \mathcal{C} . Most of these results follow in an identical manner to the original, but with small conditions on the choice of sets, and differing constant powers in the equations (largely from use of $k_{\mathcal{H}}$ rather than v_H). We include these modified results for completeness.

We write \mathcal{V}_0 to be the collection of $k_{\mathcal{H}}$ -sets of $V = V(\mathcal{H}(n, p))$, with a single vertex from each partition of $\mathcal{H}(n, p)$. For a set $Y \subseteq V$, with $|Y| \leq k_{\mathcal{H}}$ and at most one vertex from each partition, we write $\mathcal{V}_0(Y)$ for the set $\{Z \in \mathcal{V}_0 : Z \supseteq Y\}$. We then extend our earlier weight function $w = w_{\mathcal{H}(n, p)}$ to these sets Y , by setting

$$w(Y) = \sum \{w(Z) : Z \in \mathcal{V}_0(Y)\}.$$

We define our new property \mathcal{C} for $\mathcal{H}(n, p)$ as follows: $\mathcal{H}(n, p)$ satisfies \mathcal{C} if for all such Y , as defined above, with $|Y| = k_{\mathcal{H}} - 1$,

$$\max w(\mathcal{V}_0(Y)) \leq \max\{n^{-2(k_{\mathcal{H}}-1)}\Phi(\mathcal{H}(n, p)), 2\text{med } w(\mathcal{V}_0(Y))\}$$

We will then prove Lemma 5.10, by proving the two following results.

Lemma 5.11. $\Pr(\mathcal{A}\mathcal{R}\overline{\mathcal{C}}) = n^{-\omega(1)}.$

Lemma 5.12. $\Pr(\mathcal{R}\mathcal{C}\overline{\mathcal{B}}) = n^{-\omega(1)}.$

We have already demonstrated the first part of the proof of Lemma 5.11 in Section 4.2. For clarity, to follow the arguments of the original paper, we will first prove Lemma 5.12, before returning to Lemma 5.11 in the next Section. We must firstly show that

$$|\{K \in \mathcal{V}_0 : w(K) \geq \delta \max w(\mathcal{V}_0)\}| = \Omega(|\mathcal{V}_0|) \quad (14)$$

Noting that $|\mathcal{V}_0| = \Omega(n^{k_{\mathcal{H}}})$, and that this implies $\max rw(\mathcal{V}_0) = \mathcal{O}(1)$. We now need another small modification of a lemma from the graph case. We let $\psi(X) = \max w(\mathcal{V}_0(X))$ and let B be a positive number, recall that V is the vertex set of $\mathcal{H}(n, p)$, and is of size $k_{\mathcal{H}}n/v_H$, with $k_{\mathcal{H}}$ partitions of size n/v_H .

Lemma 5.13. *Suppose that for each $Y \subseteq V$, satisfying $|Y| = k_{\mathcal{H}} - 1$ and with at most one vertex from each partition of V , and $\psi(Y) \geq B$ we have*

$$\left| \left\{ Z \in \mathcal{V}_0(Y) : w(Z) \geq \frac{1}{2}\psi(Y) \right\} \right| \geq \frac{n/v_H - k_{\mathcal{H}}}{2}.$$

Then for any $X \subseteq V$ with $|X| = k_{\mathcal{H}} - i$, at most one vertex from each partition and $\psi(X) \geq 2^{i-1}B$, we have

$$\left| \left\{ Z \in \mathcal{V}_0(X) : w(X) \geq \frac{1}{2^i}\psi(X) \right\} \right| \geq \left(\frac{n/v_H - k_{\mathcal{H}}}{2} \right)^i \frac{1}{(i-1)!}. \quad (15)$$

Proof. We write N_i for the right hand side of the above equation, and proceed by induction on i . The case $i = 1$ is the hypothesis of the lemma. We then assume X as stated, and choose $Z \in \mathcal{V}_0(X)$ with $w(Z) = \psi(X)$ (i.e. Z such that w is maximal). We let $y \in Z \setminus X$ and $Y = X \cup \{y\}$. We then have that $|Y| = k_{\mathcal{H}} - (i - 1)$ and that $\psi(Y) = \psi(X) \geq 2^{i-1}B (\geq 2^{i-2}B)$, and so, by the inductive hypothesis there are at least N_{i-1} sets $Z' \in \mathcal{V}_0(Y)$ with $w(Z') \geq 2^{-(i-1)}\psi(Y)$. For each such Z' , $Z' \setminus \{y\}$ is a $(k_{\mathcal{H}} - 1)$ -subset of V with $\psi(Z' \setminus \{y\}) \geq w(Z') \geq B$. So then for each such Z' , there are at least $(n/v_H - k_{\mathcal{H}})/2$ sets $Z'' \in \mathcal{V}_0(Z' \setminus \{y\})$ with

$$w(Z'') \geq \psi(Z' \setminus \{y\})/2 \geq 2^{-i}\psi(X).$$

Therefore the number of such pairs, (Z, Z'') is at least $N_{i-1}(n/v_H - k_{\mathcal{H}})/2$. Equally, for each Z'' , each corresponding Z' is $Z'' \setminus \{u\} \cup \{y\}$ for some $u \in Z'' \setminus (X \cup \{y\})$. Therefore the number of such Z' is at most $i - 1$, providing our factorial term. This completes the proof. \square

We now continue the proof of Lemma 5.12. We set $\delta = 2^{k_{\mathcal{H}}}$ and then \mathcal{C} implies the hypothesis of the above lemma, with $B = (2n)^{-(k_{\mathcal{H}}-1)}\Phi(\mathcal{H}(n, p)) (> n^{-2(k_{\mathcal{H}}-1)}\Phi(\mathcal{H}(n, p)))$. We also clearly have that

$$\psi(\emptyset) \geq n^{-(k_{\mathcal{H}}-1)}\Phi(\mathcal{H}(n, p)) = 2^{k_{\mathcal{H}}-1}B.$$

We also set $\gamma = (2^{k_{\mathcal{H}}+1}(k_{\mathcal{H}} - 1)!)^{-1}$, and using (15), we now have

$$|\{K \in \mathcal{V}_0 : w(K) \geq \delta \max w(\mathcal{V}_0)\}| > \gamma n^{k_{\mathcal{H}}}.$$

We let J be the largest power of 2, not exceeding $\max w$ and

$$\mathcal{Z} = \{Z \in \mathcal{V}_0 : w(Z) > \delta J\}.$$

In a sense, \mathcal{Z} can be thought of as the vertex sets of size $k_{\mathcal{H}}$, whose complement has a relatively large number of factors. For any set $X \subseteq V$ with $|X| \leq k_{\mathcal{H}}$, with at most a single vertex from each partition, let $\mathcal{Z}(X) = \{Z \in \mathcal{Z} : X \subset Z\}$, and say such a set X is *good* if $|\mathcal{Z}(X)| > \gamma n^{k_{\mathcal{H}} - |X|}$. In particular we know that the empty set is *good*. We then fix an ordering $a_1, \dots, a_{k_{\mathcal{H}}}$ of $V(\mathcal{H})$. For distinct vertices, $x_1, \dots, x_r \in \mathcal{H}(n, p)$, we define $S(x_1, \dots, x_r)$ to be the collection of copies ϕ of \mathcal{H} in the complete multigraph KM_n , for which

$$\phi(a_i) = x_i \text{ for } i \in [r],$$

$$\phi(a_i)\phi(a_j) \in E(\mathcal{H}(n, p)) \text{ whenever } i, j \geq r \text{ and } a_i a_j \in E(\mathcal{H})$$

and that $\phi(V(\mathcal{H})) \in \mathcal{Z}$.

For each $r \in \{0, \dots, k_{\mathcal{H}}\}$ let $N_r = N(a_r) \cap \{a_{r+1}, \dots, a_{k_{\mathcal{H}}}\}$, and $d_r = |N_r|$, where $N(a_r)$, means the neighbourhood of a_r (in \mathcal{H}). We now let $\mathcal{Y}(x_1, \dots, x_r)$ be the event

$$\{|S(x_1, \dots, x_r)| = \Omega(p^{d_r + \dots + d_{k_{\mathcal{H}}-1} n^{k_{\mathcal{H}}-r})\}.$$

Note that in particular we have, $d_{k_{\mathcal{H}}} = 0$, and, recalling that $\mathcal{H}(G)$, is the set of copies of \mathcal{H} in $\mathcal{H}(n, p)$,

$$S(\emptyset) = \{\phi \in \mathcal{H}(G) : w(\phi(V(\mathcal{H}))) > \delta J\}$$

and that $\mathcal{Y}(\emptyset)$ is the event

$$\{|S(\emptyset)| = \Omega(p^{h_{\mathcal{H}}} n^{k_{\mathcal{H}}})\}.$$

Then, for $v_1, \dots, x_r \in V$, and from distinct partitions, let $Q(x_1, \dots, x_r)$ be the event

$$\{\{x_1, \dots, x_r\} \text{ is good}\} \wedge \overline{\mathcal{Y}}(x_1, \dots, x_r).$$

Since we have shown that \mathcal{C} implies that the empty set is good, we have that $\overline{\mathcal{B}\mathcal{C}} \subseteq Q(\emptyset)$ and therefore, we simply require to show that

$$\Pr(\mathcal{R}Q(\emptyset)) = n^{-\omega(1)},$$

to prove Lemma 5.12, we continue (as always, in the same fashion as for the graph case), by proving a slightly more general argument for induction purposes, namely, that for any choice of r and vertices x_1, \dots, x_r ,

$$\Pr(\mathcal{R}Q(x_1, \dots, x_r)) = n^{-\omega(1)}. \quad (16)$$

Our induction is on $k_{\mathcal{H}} - r$, with our initial step $r = k_{\mathcal{H}}$, trivially following since the definition of being good for subsets of size $k_{\mathcal{H}}$ is to belong to \mathcal{Z} . For general $r < k_{\mathcal{H}}$, we set $X = \{x_1, \dots, x_r\}$. and then let \mathcal{P} be the event

$$\{y \in V \setminus X, X \cup \{y\} \text{ good} \Rightarrow \mathcal{Y}(x_1, \dots, x_r, y)\}.$$

By the inductive hypothesis we know that $\Pr(\mathcal{R}\overline{\mathcal{P}}) = n^{-\omega(1)}$, so we only need to show

$$\Pr(\mathcal{R}\mathcal{P}Q(x_1, \dots, x_r)) < n^{-\omega(1)}.$$

We also note that if X is good then,

$$|y : X \cup \{y\} \text{ good}| = \Omega(n). \quad (17)$$

To ensure that the edges between x_r and $V \setminus X$ are independent of the initial conditioning, we use a relaxed form of \mathcal{R} , \mathcal{R}_X , which we say is satisfied if it satisfies part (a) of \mathcal{R} , whenever $A = \{a_1, \dots, a_r\}$, $\psi(a_i) = x_i$ ($i \in [r]$) and $E' \subseteq (H - A)$.

As for the graph case, if $\mathcal{R}\mathcal{P} \wedge \{X \text{ good}\}$ holds, but $\mathcal{Y}(x_1, \dots, x_r)$ does not, then there must be some $J = 2^{k_{\mathcal{H}}}$, with $k_{\mathcal{H}}$ and integer not exceeding $n \log n$ (the magnitude of the log of the number of \mathcal{H} -Factors in the complete multigraph), such that with \mathcal{Z} good, we have the following, (noting that throughout this chapter, wherever we choose sets of vertices from $V - X$, we choose them from partitions that do not contain vertices of X),

- (a) \mathcal{R}_X holds;
- (b) There are at least $\Omega(n)$ y 's in $V \setminus X$ for which we have $\mathcal{Y}(x_1, \dots, x_r, y)$ (by (17)), and lastly,
- (c) $\mathcal{Y}(x_1, \dots, x_r)$ does not hold.

We note, that for a given J , the first two properties, depend only on $G' := G - X$. Since the number of possibilities for J is at most $n \log n$, it is enough to show that for any J and G' satisfying (a) and (b) (with respect to J),

$$\Pr(\overline{\mathcal{Y}}(x_1, \dots, x_r) | G') = n^{-\omega(1)}.$$

Given this fixed G' , we can express $|S(x_1, \dots, x_r)|$ as a multilinear polynomial in terms of the indicator variable for edges between x_r and other vertices of $\mathcal{H}(n, p)$, i.e.

$$t_u := \mathbf{1}_{\{x_r u \in E(\mathcal{H}(n, p))\}} \quad u \in V \setminus X.$$

giving the polynomial,

$$|S(x_1, \dots, x_r)| = g(t) := \sum_U \alpha_U t_U,$$

where U ranges over d_r subsets of $V \setminus X$, with the vertices taken from the correct partitions, corresponding to the edges from $a_i = \phi^{-1}(x_i)$ in \mathcal{H} , and α_U is the number of copies ψ of $K := \mathcal{H} - \{a_1, \dots, a_r\}$ in G' with the induced subgraph,

$$\psi(N_r) = U$$

and

$$\psi(\{a_{r+1}, \dots, a_{k_{\mathcal{H}}}\}) \cup X \in \mathcal{Z}.$$

We now apply a concentration result from Section 5.1, namely Theorem 5.3. To apply this, we require a normal polynomial, so we normalise, and consider

$$f(t) = \alpha^{-1}g(t),$$

where α is the maximum of the α_U 's. The hypothesis requires that $\mathbb{E}f = \omega(\log n)$ and $\max_{1 \leq j \leq d-1} \mathbb{E}_j f \leq n^{-\epsilon} \mathbb{E}f$, and will allow us to say that it is close to expectation. We rewrite:

$$g(t) = \sum_{y \in V \setminus X} \sum \{t_{\phi(N_r)} : \phi \in S(x_1, \dots, x_r, y)\}.$$

Since we know that the indicator variables, t_u , $u \in V \setminus X$, are independent of G' , which determines our sets $S(x_1, \dots, x_r, y)$, and using property (b) our situation, we have

$$\mathbb{E}g = p^{d_r} \sum_{y \in V \setminus X} |S(x_1, \dots, x_r, y)| = \Omega(p^{d_r + \dots + d_{k_{\mathcal{H}}-1}} n^{k_{\mathcal{H}}-r}). \quad (18)$$

Noting that if $d_r = 0$, then there are no random edges to consider, and we have that $|S(x_1, \dots, x_r)|$ will equal

$$\sum_{y \in V \setminus X} |S(x_1, \dots, x_r, y)| = \Omega(p^{d_r + \dots + d_{k_{\mathcal{H}}-1}} n^{k_{\mathcal{H}}-r}),$$

as required, we will now assume that $d_r > 0$.

We now set $\mathcal{H}' = \mathcal{H} - \{a_1, \dots, a_{r-1}\}$, and so $d_r + \dots + d_{k_{\mathcal{H}}-1} = e(\mathcal{H}')$ and that $k_{\mathcal{H}} - r = v(\mathcal{H}') - 1$. This gives us that the right hand side of the expectation of g , is $\Omega(p^{e(\mathcal{H}')} n^{v(\mathcal{H}')-1})$. Using that $p > n^{-1/m(H)}$ and that \mathcal{H} contains no subgraphs of density $m(H)$, we have

$$\mathbb{E}g = \begin{cases} \omega(\log n) & \text{if } r = 1 \\ n^{\Omega(1)} & \text{if } r > 1. \end{cases}$$

We will now show that for the normalised polynomial $f(t)$, that

$$\mathbb{E}f = \omega(\log n) \quad (19)$$

and

$$\max\{\mathbb{E}_T f : T \subseteq V \setminus X, 0 < |T| < d_r\} = n^{-\Omega(1)} \mathbb{E} f. \quad (20)$$

With these conditions, the concentration theorem will tell us that f , and therefore g is close to its expectation, which implies $\overline{\mathcal{Y}}$, as required. To prove the two conditions, we will find it easier to consider the partial derivatives of g rather than f . We use $t_e = \mathbf{1}_{e \in E(\mathcal{H}(n,p))}$, $t_S = \prod_{e \in S} t_e$ and $t = (t_e : e \in E(KM_{V \setminus X}))$, where $KM_{V \setminus X}$ is the multigraph induced by the ‘complete’ multigraph KM_n on the vertex set $V \setminus X$.

Since we are only interested in establishing upper bounds on the partial derivatives of g , we may now disregard the second requirement on α_U , namely $\psi(\{a_{r+1}, \dots, a_{k_{\mathcal{H}}}\}) \cup X \in \mathcal{Z}$. Therefore we are left with

$$p^{-(d_r-l)} \mathbb{E}_T g \leq \tau(t) := \sum_{\phi} t_{\phi(E(K))},$$

where we sum over ϕ , injections such that

$$\phi : V(K) \rightarrow V \setminus X \text{ with } \phi(N_r) \supseteq T.$$

We set E^* , as before to be $E^* = \max\{\mathbb{E}_L \tau : L \subseteq E(KM_{V \setminus X}), |L| < |E(K)|\}$. We will show that there is a positive constant ε (depending only on \mathcal{H}), such that (for large enough n),

$$p^{d_r-l} \mathbb{E}^* < n^{-\varepsilon} \mathbb{E} g. \quad (21)$$

This will give us the two requirements for our concentration theorem, as follows. To prove (19), we need to show that

$$\alpha^{-1} \mathbb{E} g = \omega(\log n).$$

We apply (21) with $T = U$, a d_r -subset of V/X . We consider the two possible conditions of \mathcal{R}_X , which gives two separate cases, firstly if $\mathbb{E} g \geq n^{\varepsilon/2}$, (i.e. $\mathbb{E}^* \geq n^{-o(1)}$) then \mathcal{R}_X tells us that

$$\alpha_U = \mathbb{E}_U g < n^{\varepsilon/4} \max\{1, \mathbb{E}^*\} \leq n^{-\varepsilon/4} \mathbb{E} f.$$

In the other case we are left with $\mathbb{E}^* < n^{-\varepsilon/2}$, and hence $\mathbb{E}^* = n^{-\Omega(1)}$. We know that $\mathbb{E} g = \omega(\log n)$, and hence we choose our $\beta(n) = \omega(1)$ (from \mathcal{R}) such that $\beta(n)^{-1} \mathbb{E} g = \omega(\log n)$. Since this does not depend on our choice of U , we have (19) as required.

To prove the other requirement, we need $\mathbb{E}_T g = n^{-\Omega(1)} \mathbb{E} g$ for any T , as we defined in (20). We again apply (21), noting that we can decrease the

ε without violating the equation, and use this observation to assume that $\varepsilon < 1/m(H)$. This gives us

$$\mathbb{E}_T g < p^{d_r-l} n^{\varepsilon/2} \max\{1, \mathbb{E}^*\} \leq n^{-\varepsilon/2} \mathbb{E} g$$

as we required.

We now return to prove (21). We fix $L \subseteq E(KM_{V \setminus X})$ and let $h_l = |E(K)| - |L|$, (recalling that $K = \mathcal{H} - \{a_1, \dots, a_r\}$). We know that $\mathbb{E}_L \tau = p^{h_l} N_L$, where again N_L is the number of ϕ , as defined just before (21), with $\phi(E(K)) \supseteq L$. Each ϕ satisfies $\phi(V(K)) \supseteq I := T \cup V(L)$, where as earlier, $V(L) \supseteq V \setminus X$ is the set of vertices incident to the edges of L . We let $I = \{i_1, \dots, i_s\}$, then we have $N_L = \sum N_L(b_1, \dots, b_s)$, where (b_1, \dots, b_s) range over s -tuples of distinct elements of $V(K)$ and we sum over the number of ϕ 's as above, with $\phi(b_j) = i_j$ for each $1 \leq j \leq s$. We only have $\mathcal{O}(1)$ choices for the b_j 's and hence the result will follow if we can show that for any such choice,

$$p^{d_r-l+h_l} N_L(b_1, \dots, b_s) = n^{-\Omega(1)} \mathbb{E} g.$$

Given a fixed choice of b_i 's, let $\mathcal{H}'' = \mathcal{H}[\{a_r, b_1, \dots, b_s\}]$. We know that $N_L(b_1, \dots, b_s) < n^{k_{\mathcal{H}}-r-s} = n^{k_{\mathcal{H}}-r-(v(\mathcal{H}'')-1)}$, and that $h_l \geq l + d_{r+1} + \dots + d_{k_{\mathcal{H}}-1} - e(\mathcal{H}'')$, since $|E(K)| = d_{r+1} + \dots + d_{k_{\mathcal{H}}-1}$ and $E(\mathcal{H}'')$ contains $\phi^{-1}(L)$ and at least l edges joining a_r to $V(K)$. Therefore we have

$$p^{d_r-l-h_l} N_L(b_1, \dots, b_s) < p^{d_r+\dots+d_{k_{\mathcal{H}}-1}} n^{k_{\mathcal{H}}-r} [n^{v(\mathcal{H}'')-1} p^{e(\mathcal{H}'')}]^{-1}.$$

Noting that since $p > n^{-1/m(H)}$ and that \mathcal{H} contains no subgraphs of density $m(H)$, we have that the expression in the square brackets is $n^{\Omega(1)}$. Combined with our earlier bound on $\mathbb{E} g$, from (18), we have the bound required above and hence (21), completing the Lemma.

5.6 Proof of Lemma 5.11

We now begin the final step of the proof, namely that $\Pr(\mathcal{AR}\bar{\mathcal{C}}) = n^{-\omega(1)}$. We maintain our use of notation from the previous chapter with $G = \mathcal{H}(n, p)$, and \mathcal{R} and \mathcal{A} . In Section 4.2, we have shown that $\mathcal{AR}\bar{\mathcal{C}}$ results in the following;

There exist x and y ; vertices in the same partition set of $G = \mathcal{H}(n, p)$ and a set Y with a single vertex from each of the remaining partitions, such that there exist, $a, b \in \text{range}(w_y)(= \text{range}(w'))$, such that $J := w_y^{-1}([a, b])$ satisfies

$$|J| > \Omega(|\mathcal{H}(y, G - R)|),$$

and

$$w_y(J) > 0.7w_y(\mathcal{H}(y, G - R)) = 0.7w(R).$$

With, $J' = w_x^{-1}([a, b])$, we also have,

$$w_y(J) > 0.7w(R),$$

and

$$w_x(J') \leq w(S) < 0.5w(R).$$

All that remains is to show that the probability of this event is $n^{-\omega(1)}$.

Once again, we can express $w_y(J)$ and $w_x(J')$ as evaluations of a multilinear polynomial in variables $\{t_u : u \in W\}$, once we have conditioned on the value of $G[W]$, in the following way. Given a set $U \subseteq W$ with $|U| \leq k_{\mathcal{H}} - 1$ and at most one vertex from each partition, let G_U^* be the graph obtained from $G[W]$ by adjoining a vertex w^* say, with neighbourhood U . We let \mathcal{K}_U be the set of copies of \mathcal{H} in G_U^* , containing $\{w^* : u \in U\}$, and

$$\alpha_U = \sum \{w'(V(K) \setminus \{w^*\}) : K \in \mathcal{K}_U, w'(V(K) \setminus \{w^*\}) \in [a, b]\}.$$

Our polynomial now becomes,

$$g(t) = \sum_{U \subseteq W} \alpha_U t_U,$$

and $w_y(J)$ and $w_x(J')$ are simply g evaluated at the point $t' := \mathbf{1}_{\{z \in W : yz \in E(G)\}}$ and $t'' := \mathbf{1}_{\{z \in W : xz \in E(G)\}}$.

\mathcal{R} tells us that $D_G(y) = \Theta(n^{k_{\mathcal{H}}-1}p^{h_{\mathcal{H}}})$ and therefore, considering that at most $o(n^{k_{\mathcal{H}}-1}p^{h_{\mathcal{H}}})$ copies of H lie in G and contain y and meet R , we have $|\mathcal{H}(y, G - R)| = \Theta(n^{k_{\mathcal{H}}-1}p^{h_{\mathcal{H}}}) = \omega(\log n)$.

We know from our conditioning of J , that it satisfies

$$|J| = \Theta(n^{k_{\mathcal{H}}-1}p^{h_{\mathcal{H}}}),$$

and

$$w_y(J) = \Theta(bn^{k_{\mathcal{H}}-1}p^{h_{\mathcal{H}}}). \quad (22)$$

We now apply a concentration result, namely Corollary 5.6. To apply it, we require a bound on the expectation of f , namely, that if $Ef \leq A$, where

$$A \geq \omega(\log n) + n^\epsilon \max_{l \neq \emptyset} E'_L f.$$

then the corollary gives us $\Pr(f > (1 + \epsilon)A) = n^{-\omega(1)}$. We fix $T \subseteq W$, with $|T| = l < k_{\mathcal{H}}$, and as always, each vertex from a different partition set. For $d = l, \dots, k_{\mathcal{H}} - 1$, and t as $t = (t_e : e \in E(KM_W))$, we consider the polynomial,

$$h_d(t) = \sum_z \sum_\phi t_{\phi(E(\mathcal{H}-z))},$$

Where z ranges over vertices of \mathcal{H} of degree d and ϕ over the injections $V(H) \setminus \{z\} \rightarrow W$ with $\phi(N_z) \supseteq T$. Then we know that

$$\alpha_T \leq b h_l(t).$$

Since $h_l(t)$ is the number of such sets, and we know that their values are bounded by b , this also gives us

$$\mathbb{E}'_T g \leq b \sum_{d>l} p^{d-l} h_d(t),$$

where \mathbb{E}'_T is the non-constant part of the partial derivative.

We let

$$\mathbb{E}_d^* = \max\{E_L h_d : L \subseteq E(KM_W), |L| < h_{\mathcal{H}} - d\}.$$

As in the previous chapter, similarly to (21) we can assert that there is positive constant ϵ dependent only on \mathcal{H} , such that, for each d ,

$$p^{d-l} \mathbb{E}_d^* < n^{-\epsilon} n^{k_{\mathcal{H}}-1} p^{h_{\mathcal{H}}}.$$

This follows from the proof of (21) in the previous chapter, by noting that in our definition of h_d , there are only finitely many z , and the inner sum, is bounded by the polynomial τ , used in (21), with $r = 1$, $a_1 = z$ and hence $d_r = d$.

We are now able to apply the concentration results we require. We consider the polynomial $f = \alpha^{-1}g$, where $\alpha = \max_U \alpha_U$. Then we have, using (22),

$$f(t') = \alpha^{-1} w_y(J) = \Theta(\alpha^{-1} b n^{k_{\mathcal{H}}-1} p^{h_{\mathcal{H}}})$$

and using the same arguments as we used at the end of the previous chapter, we can show

$$f(t') = \omega(\log n)$$

and

$$\max\{\mathbb{E}'_T f : T \subseteq W, T \neq \emptyset\} = n^{-\Omega(1)} f(t').$$

For the final step, we first summarise that we have shown $\mathcal{AR}\overline{\mathcal{C}}$ implies that there exist Y , x , y and a , $b \in \text{range}(w')$ for which we have the above two conditions, and

$$f(t'') < 0.8f(t').$$

However, we know that for any given choice of Y , x , y , a , and b , f depends only on $G[W]$. Equally, if $G[W]$ is fixed, then t and t' are independent random variables, with ‘law’ $\text{Bin}(W, p)$ (i.e. $t = (t_w : w \in W)$ has ‘law’ $\text{Bin}(W, p)$ if each t_w is an independent Bernoulli with mean p .) We finish with the final claim of [6], which can be applied directly to our result here, without generalisation, and completes the proof.

Claim 1. *For any $\epsilon > 0$ and d the following holds. If f is a multilinear, normal polynomial of degree at most d in n variables, $\zeta(n) = \omega(\log n)$, and t', t'' are independent, each with law $\text{Bin}([n], p)$, then*

$$\Pr(f(t') > \max\{\zeta(n), n^\epsilon \max_{T \neq \emptyset} E'_T f, (1 + \epsilon)f(t'')\}) = n^{-\epsilon}.$$

Because there are only polynomially many possibilities for Y, x, y, a and b , this gives us Lemma 5.11.

Proof of claim. Set

$$A = \frac{1}{2} \max \left\{ \zeta(n), n^\epsilon \max_{T \neq \emptyset} E'_T f \right\}.$$

If $\mathbb{E}f \leq A$ then Corollary 5.6 gives

$$\Pr(f(t') > A) = n^{-\omega(1)};$$

otherwise, by Theorem 5.5,

$$\Pr(f(t') > (1 + \epsilon)f(t'')) < \Pr(\max\{|f(t') - \mathbb{E}f|, |f(t'') - \mathbb{E}f|\} > (\epsilon/3)\mathbb{E}f) = n^{-\omega(1)}.$$

□

6 Theorems 1.2 and 1.3

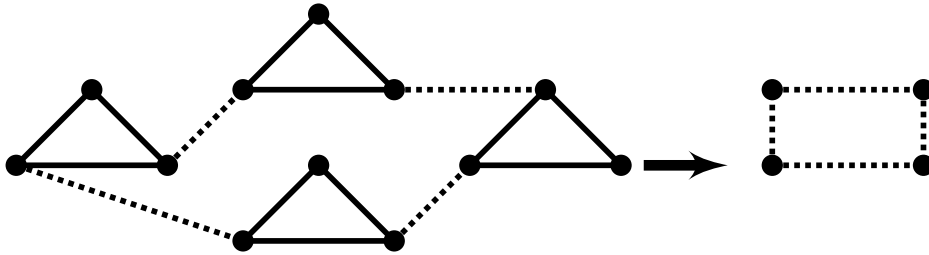
To prove Theorem 1.2, we note that our proof of Theorem 1.2 does not require the collapsed graph to be a multigraph. Throughout, we only require that the graph does not contain any subgraphs of density $m(H)$, and this follows, in our main result, from the vertex collapsing technique, but as in [6], it can also follow from strict balance of \mathcal{H} , or equivalently H . Since our partitions were fixed only by the partial embedding of subgraphs of density $m(H)$, in the strictly balanced case, we can simply choose our vertex partitions freely, and then continue with the proof.

Lastly, to prove 1.3, we use Theorem 1.2. Partition the vertices of $G(n, p)$ as usual into v_H equal sets, either as a result of vertex collapsing dense subgraphs, or freely in the strictly balanced case. In the former case, embedding the partial factors of dense directed subgraphs, requires a simple modification of Theorem 2.1, which follows the same arguments as the original but for directed graphs. A proof of which will appear in [10]. We again consider the edges between the partitions, only where they correspond to edges in H . Now however, we can consider only the edges that are in the direction we require, which are distributed uniformly and independently at random, but with edge probability $p/2$, since we discount those edges in the wrong direction. At this point, the random digraph is equivalent to $\mathcal{H}(n, p/2)$, and we apply 1.2 directly, providing the directed factor as required.

7 Conclusion

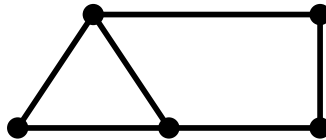
With these results, Conjecture 1 is now proven for both strictly balanced graphs, and all non-vertex balanced graphs. The methods used here also can be used to prove that the conjecture holds for a wide range of vertex-balanced H .

We can prove a threshold for a variety of ‘necklace’ and related graphs formed of copies of a dense graph linked by a ‘supergraph’ of equal or lower density.



As an example, in the above case, the threshold for finding a copy of this ‘triangle necklace’ must be at least that of a triangle factor, and Conjecture 1 suggests it should be equal. We prove this by first embedding a triangle factor, and then use 1.2 to embed the dotted edges that form a less dense cycle in the supergraph. Provided the resulting graph or multigraph from the collapsing process is less dense than $m(H)$, we will always be able to apply Theorem 1.2 to find this.

The supergraph cannot ever be denser than $m(H)$, as this implies a subgraph of density greater than $m(H)$, which is a contradiction. So, all that remains to consider are graphs where we will be left with a supergraph of equal density, after collapsing all subgraphs. We know this must be strictly balanced, or we would have continued to collapse its subgraphs and so we can apply Theorem 1.2. However, the supergraph may have fewer edges than each of the collapsed subgraphs, and certainly has fewer than the original graph. This may force it to require a higher threshold, by a constant power of log to embed, and so will not prove the conjecture.



For the above graph, the conjectured threshold is $n^{-2/3}(\log n)^{1/5}$, but using the methods within this paper, we are only able to embed it at $p =$

$\mathcal{O}(n^{-2/3}(\log n)^{1/3})$, since collapsing the triangle will leave us with another triangle, requiring the higher log term to embed.

In light of this, we see that we can prove Conjecture 1 for all H , except for those mentioned above, for which we are still within a constant power of log of the conjectured bound.

References

- [1] Noga Alon and Raphael Yuster, *Threshold functions for H -factors*, Combinatorics, Probability and Computing **2** (1993), no. 02, 137–144.
- [2] Béla Bollobás and Andrew Thomason, *Random graphs of small order*, Random graphs '83 (Poznań, 1983), North-Holland Math. Stud., vol. 118, North-Holland, Amsterdam, 1985, pp. 47–97. MR 860586 (87k:05137)
- [3] P. Erdős and A. Rényi, *On the existence of a factor of degree one of a connected random graph*, Acta Math. Acad. Sci. Hungar. **17** (1966), 359–368. MR 0200186 (34 #85)
- [4] A.M. Frieze, *Private communication*, 2012.
- [5] Svante Janson, Tomasz Łuczak, and Andrzej Ruciński, *Random graphs*, Wiley-Interscience, 2000.
- [6] Anders Johansson, Jeff Kahn, and Van Vu, *Factors in random graphs*, Random Struct. Algorithms **33** (2008), no. 1, 1–28.
- [7] Jeong Han Kim and Van Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434.
- [8] M. Krivelevich, *Embedding spanning trees in random graphs*, ArXiv e-prints (2010).
- [9] Tomasz Łuczak and Andrzej Ruciński, *Tree-matchings in graph processes*, SIAM J. Discrete Math. **4** (1991), no. 1, 107–120.
- [10] Andrew McDowell, *PhD thesis*, As yet unpublished.
- [11] Andrzej Ruciński, *Matching and covering the vertices of a random graph by copies of a given graph*, Discrete Math. **105** (1992), no. 1-3, 185–197. MR 1180202 (94a:05180)

- [12] Andrzej Ruciński and Andrew Vince, *Balanced extensions of graphs and hypergraphs*, *Combinatorica* **8** (1988), no. 3, 279–291.
- [13] Joel H. Spencer, *Threshold functions for extension statements*, *J. Comb. Theory, Ser. A* **53** (1990), no. 2, 286–305.
- [14] Van Vu, *On the concentration of multivariate polynomials with small expectation*, *Random Struct. Algorithms* **16** (2000), 344–363.
- [15] Van. Vu, *Concentration of non-Lipschitz functions and applications*, *Random Structures. Algorithms* **20** (2002), no. 3, 262–316.